Annals of Mathematics
Vol. 45, No. 4, October, 1944

QUASIQUATERNION ALGEBRAS

By A. A. Albert

(Received November 2, 1943)

1. Introduction

The algebraic theory[1] of associative central simple algebras is concerned chiefly with the automorphisms of such algebras, with their subalgebras, with the conditions that they be division algebras, and with the conditions that two differently generated algebras be isomorphic. These topics are surely of equal importance in the theory of nonassociative central simple algebras and there is, in addition, the study of the conditions that two algebras be isotopic.

A study of these topics in the nonassociative case has only been made for such algebras as Lie algebras, Jordan algebras, and Cayley algebras. There are central simple algebras which are much more like the associative crossed product algebras than are any of the above types and no study has yet been made of these topics for them. It is natural then to begin this new study with the discussion of an interesting special case.

We shall consider here certain algebras of order four over a general field $\mathfrak{F}$. They are a generalization of the "generalized quaternion" algebras of L. E. Dickson which have proved to be such an important special case in the associative theory, and specialize to the associative case when a single parameter $\tau = 0$. The algebras are not associative if and only if $\tau \neq 0, 1$ and are central simple.[2]

We shall show that the theory of isomorphism, automorphisms, proper subalgebras, and division algebras for these algebras is actually much simpler than in the associative case, a result quite contrary to what might have been expected. Indeed, the theory is remarkably like the theory of quadratic fields. We shall also obtain a complete study of the conditions that two of our quasiquaternion division algebras over a field of characteristic not two be isotopic.[3]

2. The multiplications of a crossed extension[4]

We consider an algebra $\mathfrak{Z}$ with a unity quantity e and let n be the order of $\mathfrak{Z}$ over the base field $\mathfrak{F}$. We call the linear transformations $a \to ax$ on $\mathfrak{Z}$ its *right multiplications* ρ_x and the transformations $a \to xa$ its *left multiplications* λ_x. Then x is non-singular if both ρ_x and λ_x are nonsingular, that is, are one-to-one.

[1] cf. Chapter IV of my *Structure of Algebras*, A. M. S. Colloquium Publications vol. 24, New York, 1939.

[2] Our algebras are permutation algebras and hence are central simple. See my *Nonassociative algebras II. New Simple Algebras*, these Annals vol. 43 (1942), pp. 708–723.

[3] A student of mine at the University of Chicago is engaged in a study of the arithmetics of quasiquaternion algebras and I expect to have the isotopy theory completed by another student.

[4] The study of crossed extensions is the subject of the paper referred to in footnote 2. The present section is an addition to that paper.

A crossed extension $\mathfrak{A} = (\mathfrak{Z}, \mathfrak{G}, \mathfrak{H}, \mathfrak{g})$ of $\mathfrak{Z}$ by $\mathfrak{G}$ and $\mathfrak{H}$ relative to the extension set $\mathfrak{g}$ is an algebra of order mn over $\mathfrak{F}$. It contains $\mathfrak{Z}$ as a subalgebra and has e as its unity quantity. The set $\mathfrak{G}$ is a group of m nonsingular transformations on $\mathfrak{Z}$ such that $eS \to e$ for every S of $\mathfrak{G}$. The set $\mathfrak{g} = \{g_{S,T}\}$ consists of nonsingular quantities $g_{S,T}$ of $\mathfrak{Z}$ with one such quantity for every pair S, T of $\mathfrak{G}$. The algebra $\mathfrak{A}$ is the supplementary sum of the m linear subspaces $j_S\mathfrak{Z}$ for S in $\mathfrak{G}$, and $j_I\mathfrak{Z} = \mathfrak{Z}$ where I is the identity transformation. The set $\mathfrak{H}$ is any subset of $\mathfrak{G}$ which contains I. Multiplication in $\mathfrak{A}$ is now defined by the distributive laws and by

$$(1) \qquad (j_S a)(j_T x) = j_{ST}\phi(S, T, x).$$

The function ϕ is the product of the three factors aT, x and $g_{S,T}$. The factors aT and x are multiplied in the order $(aT)x$ or $x(aT)$ according as S is or is not in $\mathfrak{H}$, and their product is multiplied on the right or left by $g_{S,T}$ according as ST is or is not in $\mathfrak{H}$. Then $(j_S a)(j_T x) = j_{ST}aR(S, T, x)$ where

$$(2) \qquad R(S, T, x) = TP_x Q_{g_{S,T}},$$

such that $P_x = \rho_x$ or λ_x according as S is or is not in $\mathfrak{H}$, $Q_g = \rho_g$ or λ_g according as ST is or is not in $\mathfrak{H}$.

The quantities a and x of the algebra $\mathfrak{A}$ may be expressed uniquely in the form

$$(3) \qquad a = \sum_S j_S a_S, \qquad x = \sum_T j_T x_T$$

for a_S and x_T in $\mathfrak{Z}$. Then

$$(4) \qquad ax = aR_x, \qquad R_x = \| R(S, T, x_T) \|$$

is an m-rowed square matrix whose element in the S-row and ST-column is given by (2).

In a similar fashion $(j_T x)(j_S a) = j_{TS}[g_{TS}(xS \cdot a)]$, or $j_{TS}[(xS \cdot a)g_{T,S}]$, or $j_{TS}[g_{T,S}(a \cdot xS)]$, or $j_{TS}[(a \cdot xS)g_{T,S}]$, where we use the product $(xS)\, a$ or $a\, (xS)$ according as T is or is not in $\mathfrak{H}$, we multiply by $g_{T,S}$ on the left or right according as TS is or is not in $\mathfrak{H}$. Then $xa \to aL_x$ where

$$(5) \qquad L_x = \| L(S, T, x_T) \|, \qquad L(S, T, x) = P_{xS} Q_{g_{T,S}},$$

and $P_x = \lambda_x$ or ρ_x according as T is or is not in $\mathfrak{H}$, $Q_g = \lambda_g$ or ρ_g according as TS is or is not in $\mathfrak{H}$.

It follows now that $\mathfrak{A}$ is a division algebra if and only if every R_x (or every L_x) with $x \neq 0$ is nonsingular.

3. Crossed extensions of commutative algebras

If $\mathfrak{Z}$ is commutative every $\rho_x = \lambda_x$ and the distinction between right and left factors disappears. This eliminates the complexity in (2) and (5) as well as the role of $\mathfrak{H}$ and we may write $\mathfrak{A} = (\mathfrak{Z}, \mathfrak{G}, \mathfrak{g})$. Then L_x (but not R_x) may be taken to have elements in $\mathfrak{Z}$ and $xa \to aL_x$ where the element in the S-row and ST-column of L_x is

$$(6) \qquad (x_T S) g_{T,S}.$$

In the special case where $\mathfrak{Z}$ is associative as well as commutative the matrix L_x has a determinant which we may call the norm of x and designate by $N(x)$. It is a quantity of $\mathfrak{Z}$. When $\mathfrak{Z}$ is a field the algebra $\mathfrak{A}$ is a division algebra if and only if $N(x) \neq 0$ for every $x \neq 0$.

We may now define a type of algebra which we shall call a *quasicyclic* algebra. We take $\mathfrak{Z}$ to be any commutative associative algebra of order n over $\mathfrak{F}$, and with a unity quantity e, and let $\mathfrak{G}$ be a cyclic group of order n. Assume that the extension set $\mathfrak{g}$ is such that $j_T = j^i$ for every $T = S^i$ where S generates $\mathfrak{G}$ and thus $j = j_S$. Assume also that $j^n = \gamma e$ for $j \neq 0$ in $\mathfrak{F}$ and then have

(7) $$(j^k x_k)(j^i a_i) = j^{k+i} a_i x_k^{(i)},\ j^{k+i-n} \gamma a_i x_k^{(i)}$$

according as $k + i < n$ or $k + i \geqq n$, where $x_k^{(i)} = x_k S^i$. The general quantity of $\mathfrak{A}$ is $x = x_1 + jx_2 + \cdots + j^{n-1}x_n$ and

(8) $$L_x = \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_{n-1} & x_n \\ \gamma x_n' & x_1' & x_2' & \cdots & x_{n-2}' & x_{n-1}' \\ \gamma x_{n-1}'' & \gamma x_n'' & x_1'' & \cdots & x_{n-3}'' & x_{n-2}'' \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \gamma x_2^{(n-1)} & \gamma x_3^{(n-1)} & \gamma x_4^{(n-1)} & \cdots & \gamma x_n^{(n-1)} & x_1^{(n-1)} \end{pmatrix}.$$

We shall call n the *construction degree* of the quasicyclic algebra $\mathfrak{A}$ and shall write $\mathfrak{A} = (\mathfrak{Z}, S, \gamma)$. Such algebras exist for every $\mathfrak{Z}$ and γ. Indeed a basis $e_1, \cdots, e_n$ of $\mathfrak{Z}$ over may always be selected (in infinitely many ways if $\mathfrak{F}$ is infinite) such that $e = e_1 + \cdots + e_n$. Then the transformation S defined by $\alpha_1 e_1 + \cdots + \alpha_n e_n \to \alpha_1 e_2 + \cdots + \alpha_{n-1} e_n + \alpha_n e_1$ has order n and leaves e unaltered.

4. Quasiquaternion algebras

A quasicyclic algebra of construction degree two is a linear space

$$\mathfrak{A} = (\mathfrak{Z}, S, \gamma) = \mathfrak{Z} + j\mathfrak{Z}$$

of order four over $\mathfrak{F}$ such that $\gamma \neq 0$ is in $\mathfrak{F}$,

(9) $$(y + jz)(a + jb) = ay + \gamma bz' + j(az + by')$$

for all a, b, y, z of the algebra $\mathfrak{Z}$ of order two over $\mathfrak{F}$. We shall call all such algebras *quasiquaternion* algebras, and shall say that $\mathfrak{A}$ is a *split* or a *nonsplit* algebra according as $\gamma = \delta^2$ or $\gamma \neq \delta^2$ for δ in $\mathfrak{F}$.

Let us now determine all possible transformations $a \to a' = aS$ defining quasiquaternion algebras. We observe first that any $\mathfrak{Z}$ of order two with a unity quantity is necessarily both a commutative and an associative algebra. Indeed $\mathfrak{Z} = e\mathfrak{F} + u\mathfrak{F}$ such that

(10) $$u^2 = \rho u + \sigma e \qquad (\rho, \sigma \text{ in } \mathfrak{F}).$$

When the characteristic of $\mathfrak{F}$ is not two we may always choose u so as to take $\rho = 0$ and we shall do this. When the characteristic of $\mathfrak{F}$ is two then $\rho = 0$ or we may choose u so as to take $\rho = 1$. We shall thus assume that $\rho = 0$ or 1 in this case.

The linear transformation

$$a = \alpha_1 e + \alpha_2 u \to \alpha_1 e + \alpha_2(\rho e - u) \tag{11}$$

and the identity transformation are the only automorphisms of $\mathfrak{Z}$ over $\mathfrak{F}$. Also (11) has order two unless $\mathfrak{F}$ has characteristic two and $\rho = 0$. The quasiquaternion algebra $\mathfrak{A}$ defined by taking S to be either the identity automorphism or the automorphism (11) is associative and it is well known that $\mathfrak{A}$ is not associative for all other values of S. We seek to determine all such transformations.

Since $e' = e$ we have $u' = \tau e + \delta u$, $(u')' = \tau e + \delta(\tau e + \delta u) = \tau(1 + \delta)e + \delta^2 u = u$. Then

$$\tau(1 + \delta) = 0, \qquad \delta^2 = 1. \tag{12}$$

The transformation given by $\delta = -1$ is

$$a = \alpha_1 + \alpha_2 u \to a' = \alpha_1 e + \alpha_2(\tau e - u). \tag{13}$$

But $\delta^2 = 1$ implies that $\delta \neq -1$ only when $\delta = 1$, the characteristic of $\mathfrak{F}$ is not two, $2\tau = 0$, $\tau = 0$, $u' = u$, S is the identity transformation. Hence S is given by (13) for

$$\tau \neq 0, \rho. \tag{14}$$

That $\tau \neq 0$ follows from $\rho = 0$ when $\mathfrak{F}$ has characteristic not two and from the fact that (13) is the identity transformation when $\tau = 0$ and $\mathfrak{F}$ has characteristic two.

5. Elementary properties

The algebra $\mathfrak{Z} = \mathfrak{F}[u]$ with unity quantity e is equal to $\mathfrak{F}[i]$ for $i = \tau^{-1}u$. Then

$$i' = \tau^{-1}(\tau e - u) = e - i, \qquad i^2 = \beta i + \alpha e \tag{15}$$

for α and β in $\mathfrak{F}$ and $\beta \neq 1$. Here $\beta = 0$ in case the characteristic of $\mathfrak{F}$ is not two. With this generation of $\mathfrak{Z}$ the linear transformation $a \to a'$ is uniquely defined by

$$a = \alpha_1 e + \alpha_2 i \to \alpha_1 e + \alpha_2(e - i) \tag{16}$$

for every α_1 and α_2 of $\mathfrak{F}$.

LEMMA 1. *Let a be in $\mathfrak{Z}$. Then $a' = a$ if and only if $a = \xi e$; $a' = -a$ if and only if $a = \xi v$; $a' = e - a$ if and only if $a = i + \xi v$, where $v = e - 2i$ and ξ is in $\mathfrak{F}$.*

For $a = a'$ if and only if $\alpha_1 = \alpha_1 + \alpha_2$, $\alpha_2 = 0$. Also

$$a + a' = (2\alpha_1 + \alpha_2)e \tag{17}$$

is zero if and only if $\alpha_2 = -2\alpha_1$, $a = \alpha_1(e - 2i)$. Finally $a + a' = e$ is equivalent to $2\alpha_1 + \alpha_2 = 1$, $\alpha_2 = 1 - 2\alpha_1$, $a = i + \alpha_1(e - 2i)$.

The quantity $v = e$ if the characteristic of $\mathfrak{F}$ is two. Otherwise $v(e + 2i) = (1 - 4\alpha)e$ and v has an inverse in $\mathfrak{Z}$ if and only if

$$1 - 4\alpha \neq 0. \tag{18}$$

Define a linear transformation on $\mathfrak{Z}$ by

$$a = \alpha_1 e + \alpha_2 i \to a^* = \alpha_1 e + \alpha_2(-e - i). \tag{19}$$

Then if (18) holds every a of $\mathfrak{Z}$ determines a unique $d = v^{-1}a$ and conversely $a = vd$ is uniquely determined by d. We then have

LEMMA 2. *Let $d = v^{-1}a$. Then $d^* = (v')^{-1}a'$.*

For $a = dv = (\delta_1 + \delta_2 i)(e - 2i) = (\delta_1 - 2\delta_2\alpha)e + (\delta_2 - 2\delta_1)i$, $a' = (\delta_1 - 2\delta_2\alpha + \delta_2 - 2\delta_1)e - (\delta_2 - 2\delta_1)i = (\delta_2 - \delta_1 - 2\delta_2\alpha)e - (\delta_2 - 2\delta_1)i$. Then $v'd^* = -(e - 2i)(\delta_1 e - \delta_2 e - \delta_2 i) = \delta_2 e - \delta_1 e - 2\delta_2\alpha e + (2\delta_1 - 2\delta_2 + \delta_2)i$ as desired.

We shall use these lemmas in the study of isotopy. Let us observe next that the mapping

$$x = y + jz \to xA = y - jz \tag{20}$$

is an automorphism of $\mathfrak{A}$. This follows from the fact that

$$(y + j_0 z)(a + j_0 b) = ay + \lambda^2\gamma bz' + j_0(az + by') \tag{21}$$

for every $j_0 = \lambda j$ where $\lambda \neq 0$ in $\mathfrak{F}$. Then (21) with $\lambda = -1$ implies that A is an automorphism of $\mathfrak{A}$. When $1 - 4\alpha \neq 0$ the quantity v has an inverse $(1 + 2i)(1 - 4\alpha)^{-1}$ and $(vx)v^{-1} = v(xv^{-1}) = xA$. Thus A may be called an *inner* automorphism in this case.

The automorphism A is the identity automorphism if $\mathfrak{F}$ has characteristic two. Write

$$i_0 = i + \mu e \tag{22}$$

and have

$$i_0' = \mu e + (e - i) = e - (\mu e + i) = e - i_0, \tag{23}$$

as well as

$$i_0^2 = \mu^2 e + \beta i + \alpha e = \beta i_0 + (\alpha + \beta\mu - \mu^2)e.$$

In particular $i_0 = \beta e - i$ defines a mapping B given by

$$\eta_1 e + \eta_2 i + \zeta_1 j + \zeta_2 ji \to \eta_1 e + \eta_2 i_0 + \zeta_1 j + \zeta_2 ji_0 \tag{24}$$

such that $[(\eta_1 e + \eta_2 i)']B = [(\eta_1 e + \eta_2 i)B]'$, and B is an automorphism of $\mathfrak{A}$ over $\mathfrak{F}$ characteristic two.

We shall write $\mathfrak{A} = (\mathfrak{Z}, S, \gamma) = (\alpha, \gamma)$ when $\mathfrak{F}$ has characteristic not two. By (21) we have

$$(\alpha, \gamma) = (\alpha, \lambda^2\gamma). \tag{25}$$

When the characteristic of $\mathfrak{F}$ is not two we may always choose u so as to take $\rho = 0$ and we shall do this. When the characteristic of $\mathfrak{F}$ is two then $\rho = 0$ or we may choose u so as to take $\rho = 1$. We shall thus assume that $\rho = 0$ or 1 in this case.

The linear transformation

$$a = \alpha_1 e + \alpha_2 u \to \alpha_1 e + \alpha_2(\rho e - u) \tag{11}$$

and the identity transformation are the only automorphisms of $\mathfrak{Z}$ over $\mathfrak{F}$. Also (11) has order two unless $\mathfrak{F}$ has characteristic two and $\rho = 0$. The quasiquaternion algebra $\mathfrak{A}$ defined by taking S to be either the identity automorphism or the automorphism (11) is associative and it is well known that $\mathfrak{A}$ is not associative for all other values of S. We seek to determine all such transformations.

Since $e' = e$ we have $u' = \tau e + \delta u$, $(u')' = \tau e + \delta(\tau e + \delta u) = \tau(1 + \delta)e + \delta^2 u = u$. Then

$$\tau(1 + \delta) = 0, \qquad \delta^2 = 1. \tag{12}$$

The transformation given by $\delta = -1$ is

$$a = \alpha_1 + \alpha_2 u \to a' = \alpha_1 e + \alpha_2(\tau e - u). \tag{13}$$

But $\delta^2 = 1$ implies that $\delta \neq -1$ only when $\delta = 1$, the characteristic of $\mathfrak{F}$ is not two, $2\tau = 0$, $\tau = 0$, $u' = u$, S is the identity transformation. Hence S is given by (13) for

$$\tau \neq 0, \rho. \tag{14}$$

That $\tau \neq 0$ follows from $\rho = 0$ when $\mathfrak{F}$ has characteristic not two and from the fact that (13) is the identity transformation when $\tau = 0$ and $\mathfrak{F}$ has characteristic two.

5. Elementary properties

The algebra $\mathfrak{Z} = \mathfrak{F}[u]$ with unity quantity e is equal to $\mathfrak{F}[i]$ for $i = \tau^{-1}u$. Then

$$i' = \tau^{-1}(\tau e - u) = e - i, \qquad i^2 = \beta i + \alpha e \tag{15}$$

for α and β in $\mathfrak{F}$ and $\beta \neq 1$. Here $\beta = 0$ in case the characteristic of $\mathfrak{F}$ is not two. With this generation of $\mathfrak{Z}$ the linear transformation $a \to a'$ is uniquely defined by

$$a = \alpha_1 e + \alpha_2 i \to \alpha_1 e + \alpha_2(e - i) \tag{16}$$

for every α_1 and α_2 of $\mathfrak{F}$.

LEMMA 1. *Let a be in $\mathfrak{Z}$. Then $a' = a$ if and only if $a = \xi e$; $a' = -a$ if and only if $a = \xi v$; $a' = e - a$ if and only if $a = i + \xi v$, where $v = e - 2i$ and ξ is in $\mathfrak{F}$.*

For $a = a'$ if and only if $\alpha_1 = \alpha_1 + \alpha_2$, $\alpha_2 = 0$. Also

$$a + a' = (2\alpha_1 + \alpha_2)e \tag{17}$$

is zero if and only if $\alpha_2 = -2\alpha_1$, $a = \alpha_1(e - 2i)$. Finally $a + a' = e$ is equivalent to $2\alpha_1 + \alpha_2 = 1$, $\alpha_2 = 1 - 2\alpha_1$, $a = i + \alpha_1(e - 2i)$.

The quantity $v = e$ if the characteristic of $\mathfrak{F}$ is two. Otherwise $v(e + 2i) = (1 - 4\alpha)e$ and v has an inverse in $\mathfrak{Z}$ if and only if

$$1 - 4\alpha \neq 0. \tag{18}$$

Define a linear transformation on $\mathfrak{Z}$ by

$$a = \alpha_1 e + \alpha_2 i \rightarrow a^* = \alpha_1 e + \alpha_2(-e - i). \tag{19}$$

Then if (18) holds every a of $\mathfrak{Z}$ determines a unique $d = v^{-1}a$ and conversely $a = vd$ is uniquely determined by d. We then have

LEMMA 2. *Let* $d = v^{-1}a$. *Then* $d^* = (v')^{-1}a'$.

For $a = dv = (\delta_1 + \delta_2 i)(e - 2i) = (\delta_1 - 2\delta_2\alpha)e + (\delta_2 - 2\delta_1)i$, $a' = (\delta_1 - 2\delta_2\alpha + \delta_2 - 2\delta_1)e - (\delta_2 - 2\delta_1)i = (\delta_2 - \delta_1 - 2\delta_2\alpha)e - (\delta_2 - 2\delta_1)i$. Then $v'd^* = -(e - 2i)(\delta_1 e - \delta_2 e - \delta_2 i) = \delta_2 e - \delta_1 e - 2\delta_2\alpha e + (2\delta_1 - 2\delta_2 + \delta_2)i$ as desired.

We shall use these lemmas in the study of isotopy. Let us observe next that the mapping

$$x = y + jz \rightarrow xA = y - jz \tag{20}$$

is an automorphism of $\mathfrak{A}$. This follows from the fact that

$$(y + j_0 z)(a + j_0 b) = ay + \lambda^2\gamma bz' + j_0(az + by') \tag{21}$$

for every $j_0 = \lambda j$ where $\lambda \neq 0$ in $\mathfrak{F}$. Then (21) with $\lambda = -1$ implies that A is an automorphism of $\mathfrak{A}$. When $1 - 4\alpha \neq 0$ the quantity v has an inverse $(1 + 2i)(1 - 4\alpha)^{-1}$ and $(vx)v^{-1} = v(xv^{-1}) = xA$. Thus A may be called an *inner* automorphism in this case.

The automorphism A is the identity automorphism if $\mathfrak{F}$ has characteristic two. Write

$$i_0 = i + \mu e \tag{22}$$

and have

$$i_0' = \mu e + (e - i) = e - (\mu e + i) = e - i_0, \tag{23}$$

as well as

$$i_0^2 = \mu^2 e + \beta i + \alpha e = \beta i_0 + (\alpha + \beta\mu - \mu^2)e.$$

In particular $i_0 = \beta e - i$ defines a mapping B given by

$$\eta_1 e + \eta_2 i + \zeta_1 j + \zeta_2 ji \rightarrow \eta_1 e + \eta_2 i_0 + \zeta_1 j + \zeta_2 j i_0 \tag{24}$$

such that $[(\eta_1 e + \eta_2 i)']B = [(\eta_1 e + \eta_2 i)B]'$, and B is an automorphism of $\mathfrak{A}$ over $\mathfrak{F}$ characteristic two.

We shall write $\mathfrak{A} = (\mathfrak{Z}, S, \gamma) = (\alpha, \gamma)$ when $\mathfrak{F}$ has characteristic not two. By (21) we have

$$(\alpha, \gamma) = (\alpha, \lambda^2\gamma). \tag{25}$$

When $\mathfrak{F}$ has characteristic two we shall write $\mathfrak{A} = (\alpha, \beta, \gamma)$ and have

$$(26) \qquad (\alpha, \beta, \gamma) = (\alpha + \beta\mu - \mu^2, \beta, \lambda^2\gamma)$$

as a consequence of (24) and (23) if $\beta \neq 0$. If $\beta = 0$ then $i_0 = \mu e + i$, $i_0^2 = (\alpha + \mu^2)e$, $j_0 = \lambda j$ yields (26) for $\beta = 0$.

6. The norm and trace functions

The general quantity of $\mathfrak{A}$ is given by

$$x = y + jz, \qquad y = \eta_1 e + \eta_2 i, \qquad z = \zeta_1 e + \zeta_1 i$$

for η_1, η_2, ζ_1 and ζ_2 in $\mathfrak{F}$. Then y and z are in $\mathfrak{Z}$ and we define

$$(27) \qquad T(x)e = y + y', \qquad N(x) = yy' - \gamma zz'.$$

We compute x^2 by the use of (19) to obtain $x^2 = y^2 + \gamma zz' + jz(y + y') = y^2 + \gamma zz' + (x - y)(y + y')$. Then $x^2 - T(x)x = \gamma zz' - yy'$, and

$$(28) \qquad x^2 = T(x)x - N(x).$$

By (17)

$$(29) \qquad T(x) = 2\eta_1 + \eta_2 .$$

Also $yy' = (\eta_1 e + \eta_2 i)(\eta_1 e + \eta_2 e - \eta_2 i) = (\eta_1^2 + \eta_1\eta_2)e + \eta_2^2 i - \eta_2^2 i^2$. Using (15) we have

$$(30) \qquad N(x) = N_1(x)e + N_2(x)i,$$

where

$$(31) \qquad N_1(x) = \eta_1^2 + \eta_1\eta_2 - \eta_2^2\alpha - \gamma(\zeta_1^2 + \zeta_1\zeta_2 - \zeta_2^2\alpha),$$

$$(32) \qquad N_2(x) = (1 - \beta)(\eta_2^2 - \gamma\zeta_2^2).$$

The trace function is in $\mathfrak{F}[e]$ for every x and the norm function $N(x)$ is in $\mathfrak{Z}$. Then $N(x)$ is a scalar multiple of e if and only if $\eta_2^2 - \gamma\zeta_2^2 = 0$. Now $x^2 - T(x)x$ is in the algebra $\mathfrak{F}[x]$ generated by x and this algebra contains $N(x)$. If $N_2(x) \neq 0$ the algebra $\mathfrak{F}[i]$ is contained in $\mathfrak{F}[x]$. When $\mathfrak{F}[x]$ is quadratic we have $\mathfrak{F}[i] = \mathfrak{F}[x]$. This proves

LEMMA 3. *Let $\mathfrak{F}[x]$ be a quadratic subalgebra of $\mathfrak{A}$ where x is not in $\mathfrak{Z}$. Then* $N_2(x) = 0$.

We have called a quasiquaternion algebra a split or nonsplit algebra according as γ is or is not the square of an element of $\mathfrak{F}$. By (30) the quantity $N(x)$ is in $\mathfrak{F}[e]$ if and only if $\eta_2^2 - \gamma\zeta_2^2 = 0$. Then if $\mathfrak{A}$ is a nonsplit algebra we have $\eta_2^2 = \gamma\zeta_2^2$ only if $\eta_2 = \zeta_2 = 0$, x is in $\mathfrak{F}[j]$. We have proved

LEMMA 4. *If $\mathfrak{A}$ is nonsplit and x is not in $\mathfrak{Z}$ the quantity $N(x)$ is in $\mathfrak{F}$ if and only if x is in $\mathfrak{F}[j]$.*

As a consequence of Lemmas 3 and 4 we have

THEOREM 1. *The only quadratic subalgebras of a nonsplit quasiquaternion algebra are its subalgebras $\mathfrak{Z}$ and $\mathfrak{F}[j]$.*

7. The left multiplications of $\mathfrak{A}$

The left multiplication L_x of a quantity $x = y + jz$ of a quasiquaternion algebra $\mathfrak{A}$ is a linear transformation on $\mathfrak{A}$ which has been exhibited[5] as

$$L_x = \begin{pmatrix} y & z \\ \gamma z' & y' \end{pmatrix}. \tag{33}$$

No difficulty arises if we regard the elements of L_x as quantities of $\mathfrak{Z}$ when we are studying a single algebra. However when we study the isotopism of quasiquaternion algebras we will be forced to consider them as linear spaces of order four over $\mathfrak{F}$ and to replace e by the two-rowed identity matrix, i by the two rowed matrix

$$\begin{pmatrix} 0 & 1 \\ \alpha & \beta \end{pmatrix}, \tag{34}$$

in the quantities y, y', z, z' which appear in (33).

The multiplication

$$L_j = \begin{pmatrix} 0 & e \\ \gamma e & 0 \end{pmatrix} \tag{35}$$

has the property

$$L_jL_x = \begin{pmatrix} \gamma z' & y' \\ \gamma y & \gamma z \end{pmatrix}, \qquad L_xL_j = \begin{pmatrix} \gamma z & y \\ \gamma y' & \gamma z' \end{pmatrix}. \tag{36}$$

But $jx = j(y + jz) = \gamma z + jy$, $xj = (y + jz)j = jy' + \gamma z'$,

$$L_jL_x = L_{xj}, \qquad L_xL_j = L_{jx}. \tag{37}$$

It follows that

$$(jx)c = j(xc), \qquad (xj)c = x(jc) \tag{38}$$

for every c and x of $\mathfrak{A}$. Thus $\mathfrak{A}$ has a partial associative law.

We also compute the sum of the matrices of (36) to obtain

$$L_xL_j + L_jL_x = L_q, \tag{39}$$

where

$$q = xj + jx = \gamma T(z)e + T(y)j, \tag{40}$$

and $T(z) = 2\zeta_1 + \zeta_2$, $T(y) = 2\eta_1 + \eta_2$ are in $\mathfrak{F}$. Then q is in $\mathfrak{F}[j]$ and L_q is in $\mathfrak{F}[L_j]$.

Let us now suppose that $c = a + jb$ is a quantity of $\mathfrak{A}$ such that L_c is nonsingular.

---

[5] Note that this is formally the same as the representation of an associative quaternion algebra which has been presented in many places. However $z \to z'$ is an automorphism in the associative case and is *not* an automorphism here.

Then

$$N(c) = aa' - bb' \tag{41}$$

is a nonsingular quantity of $\mathfrak{Z}$. Write $d = [N(c)]^{-1}$ and have

$$L_c^{-1} = \begin{pmatrix} da' & -db \\ -\gamma db' & da \end{pmatrix}, \tag{42}$$

$$L_j L_c^{-1} + L_c^{-1} L_j = \begin{pmatrix} -\gamma b'd & ad \\ \gamma a'd & -\gamma bd \end{pmatrix} + \begin{pmatrix} -\gamma bd & a'd \\ \gamma ad & -\gamma b'd \end{pmatrix}. \tag{43}$$

It follows that $L_j L_c^{-1} + L_c^{-1} L_j = \sigma_1 I + \sigma_2 L_j$ for σ_1 and σ_2 in $\mathfrak{F}$ if and only if

$$-\gamma(b + b') = \sigma_1 N(c), \qquad a + a' = \sigma_2 N(c). \tag{44}$$

However by (17) we see that either $\sigma_1 = \sigma_2 = 0$ or $N(c)$ is in $\mathfrak{F}$. By Lemma 3 we see that if $\mathfrak{A}$ is nonsplit either c is in $\mathfrak{F}[j]$ or $b + b' = a + a' = 0$. Applying Lemma 1 we have $a = \xi v$, $b - \eta v$, $c = \xi v + \eta j v = v(\xi e - \eta j) = vg$. We have proved

Theorem 2. *Let $\mathfrak{A}$ be a nonsplit quasiquaternion algebra and c be a left nonsingular quantity of $\mathfrak{A}$ such that $L_j L_c^{-1} + L_c^{-1} L_j$ is in $\mathfrak{F}[L_j]$. Then c is either in $\mathfrak{F}[j]$ or has the form $c = vg$ where g is in $\mathfrak{F}[j]$,*

$$L_c = \begin{pmatrix} \xi v & \eta v \\ \gamma \eta v' & \xi v' \end{pmatrix} = \begin{pmatrix} \xi e & -\eta e \\ -\gamma \eta e & \xi e \end{pmatrix} \begin{pmatrix} v & 0 \\ 0 & v' \end{pmatrix} = L_g L_v . \tag{45}$$

If $c = \eta_1 e + \eta_2 j$ for η_1 and η_2 in $\mathfrak{F}$ and not both zero then L_c is nonsingular. By (37) we have $L_c L_x = (\eta_1 I + \eta_2 L_j) L_x = L_{xc}$, $L_x = L_c^{-1} L_{xc}$. The linear space $L(\mathfrak{A})$ of all left multiplications of $\mathfrak{A}$ has order four and so does $L_c^{-1} L(\mathfrak{A})$. But $L_c^{-1} L(\mathfrak{A})$ contains $L(\mathfrak{A})$, $L_c^{-1} L(\mathfrak{A}) = L(\mathfrak{A})$.

Suppose next that $c = vg$ where g is in $\mathfrak{F}[j]$. Then if v is a nonsingular quantity of $\mathfrak{Z}$ we have $L_c^{-1} = (L_g^{-1} L_v)^{-1} = L_v^{-1} L_g^{-1}$, $L_c^{-1} L(\mathfrak{A}) = L_v^{-1} L(\mathfrak{A})$. But by Lemma 2 we have

$$L_v^{-1} L_x = \begin{pmatrix} v^{-1} & 0 \\ 0 & v'^{-1} \end{pmatrix} \begin{pmatrix} y & z \\ \gamma z' & y' \end{pmatrix} = \begin{pmatrix} y_0 & z_0 \\ \gamma z_0^* & y_0^* \end{pmatrix},$$

where $y = v y_0$, $z = v z_0$. Thus $L_v^{-1} L_x$ is the left multiplication of the general quantity of the algebra $\mathfrak{A}_0$ with a basis $1, i, j, ji$ such that $i^2 = \beta i + \alpha e$, $j^2 = \gamma e$, $ij = j(-e - i)$. This algebra is $\mathfrak{A}$ if $\mathfrak{F}$ has characteristic two. Otherwise it is isomorphic to $\mathfrak{A}$ since in $\mathfrak{A}_0$ we have $i_0^2 = \alpha e$, $i_0 j = j(e - i_0)$ if $i_0 = -i$. Then the linear transformation replacing i by $-i$ replaces $L_v^{-1} L_x$ by $P L_{x_0} P^{-1}$ and we have proved

Theorem 3. *Let $\mathfrak{A}$ be a nonsplit algebra, c be a left nonsingular quantity of $\mathfrak{A}$ such that $L_j L_c^{-1} + L_c^{-1} L_j$ is in $\mathfrak{F}[L_j]$. Then $L_c^{-1} L(\mathfrak{A}) = P L(\mathfrak{A}) P^{-1}$ for a nonsingular linear transformation P.*

8. Quadratic subalgebras

The results we shall obtain on the isomorphisms of quasiquaternion algebras will be consequences of a determination of their quadratic subalgebras. We

have already made this determination for nonsplit algebras in Theorem 1. Hence we may assume that $\mathfrak{A}$ is a split algebra. By (25) and (26) we may take $\gamma = 1$. Let $\mathfrak{F}[x]$ be a quadratic subalgebra of $\mathfrak{A}$ and $\mathfrak{F}[x] \neq \mathfrak{Z}$. By Lemma 3 we have $\eta_2^2 = \zeta_2^2$, $\zeta_2 = \epsilon\eta_2$ where $\epsilon = \pm 1$. Then $N(x) = N_1(x) = (\eta_1^2 + \eta_1\eta_2 - \eta_2^2\alpha - \zeta_1^2 - \epsilon\zeta_1\eta_2 + \eta_2^2\alpha)e = [\eta_1^2 - \zeta_1^2 + (\eta_1 - \epsilon\zeta_1)\eta_2]e$ and thus

$$N(x) = (\eta_1 - \epsilon\zeta_1)(\eta_1 + \epsilon\zeta_1 + \eta_2)e. \tag{46}$$

We study first the case where $T(x) = 2\eta_1 + \eta_2 = 0$. Then $\eta_1 + \eta_2 = -\eta_1$ and

$$x^2 = -N(x) = (\eta_1 - \epsilon\zeta_1)^2. \tag{47}$$

If $\eta_1 = \epsilon\zeta_1$ then $\zeta_1 = \epsilon\eta_1$, $\eta_2 = -2\eta_1$, $\zeta_2 = -2\epsilon\eta_1$ and $x^2 = 0$. We have proved

THEOREM 4. *A quantity x in $\mathfrak{A} = (\mathfrak{Z}, S, 1)$ which is not in $\mathfrak{Z}$ has the property that $x^2 = 0$ if and only if $x = \eta w$ or $\eta(wA)$ where $\eta \neq 0$ is in $\mathfrak{F}$ and*

$$w = (e + j)v, \qquad wA = (e - j)v. \tag{48}$$

Suppose next that $x^2 \neq 0$ but $T(x) = 0$. Then by (47) we may normalize x so that $x^2 = e$, $\eta_1 - \epsilon\zeta_1 = \mp 1$, $\zeta_1 = \epsilon\eta_1 \mp 1$, $\zeta_2 = \epsilon\eta_2$, $\eta_2 = -2\eta_1$, $x = \eta_1 w \pm j$ or $\eta_1(wA) \pm j$.

THEOREM 5. *The quantities*

$$x = \eta w \pm j, \qquad \eta(wA) \pm j \tag{49}$$

of $\mathfrak{A} = (\mathfrak{Z}, S, 1)$ defined for η in $\mathfrak{F}$ and w of (45) *have the property that $x^2 = e$. They are the only quantities of $\mathfrak{A}$ with this property.*

A quadratic algebra $\mathfrak{F}[x]$ of characteristic not two has a generation with $T(x) = 0$. Thus we have

THEOREM 6. *Let $\mathfrak{A} = (\alpha, 1)$ over $\mathfrak{F}$ of characteristic not two. Then the quadratic subalgebras of $\mathfrak{A}$ are $\mathfrak{Z}$, $\mathfrak{F}[w]$, $\mathfrak{F}[wA]$ and the algebras $\mathfrak{F}[x]$ such that $x^2 = e$ defined by* (49).

If the characteristic of $\mathfrak{F}$ is two the quantity $v = 1$ and (48), (49) imply that the only quadratic subalgebra $\mathfrak{F}[x]$ such that $T(x) = 0$ and $\mathfrak{F}[x] \neq \mathfrak{Z}$ is $\mathfrak{F}[j]$. There remain the subalgebras $\mathfrak{F}[x]$ with $T(x) \neq 0$. Then we may assume that $T(x) = 1 = 2\eta_1 + \eta_2 = \eta_2$, $N(x) = (\eta_1 + \zeta_1)(\eta_1 + \zeta_1 + 1)e$, $x^2 = x + (\eta_1^2 + \zeta_1^2 + \eta_1 + \zeta_1)e$. Put $x_0 = x - (\eta_1 + \zeta_1)e$ and see that

$$\mathfrak{F}[x_0] = \mathfrak{F}[x], \qquad x_0^2 = x^2 - (\eta_1^2 + \zeta_1^2)e = x + (\eta_1 + \zeta_1)e = x_0 .$$

Hence we may assume that $N(x) = 0$, and thus that $\eta_1 + \zeta_1 = 0$ or that $\eta_1 + \zeta_1 = 1$. In the former case $\eta_1 = \zeta_1$, $\zeta_2 = \eta_2 = 1$ and $x = \eta_1(e + j) + i + ji$. In the latter case $x = e + \eta_1(e + j) + i + ji$ which generates the same algebra as does $\eta_1(e + j) + i + ji$. We have proved

THEOREM 7. *Let $\mathfrak{A}$ be a split quasiquaternion algebra over $\mathfrak{F}$ of characteristic two. Then the only quadratic subalgebras of $\mathfrak{A}$ other than $\mathfrak{Z}$ and $\mathfrak{F}[j]$ are the algebras $\mathfrak{F}[x]$ such that $x^2 = x$ where x is defined by*

$$(50) \qquad x = (e + j)\eta + i + ji,$$

for η in $\mathfrak{F}$.

Note that if $\mathfrak{A} = (\alpha, 0, 1)$ there are two subalgebras of $\mathfrak{A}$ containing quantities which are the roots of inseparable quadratic equations, but if $\mathfrak{A} = (\alpha, \beta, 1)$ with $\beta \neq 0$ there is only one such algebra. Hence no $(\alpha, 0, 1)$ is isomorphic to an $(\alpha, \beta, 1)$ with $\beta \neq 0$.

9. Generations of quasiquaternion algebras

We seek all generations $\mathfrak{A} = \mathfrak{Z}_0 + j_0\mathfrak{Z}_0$ of a quasiquaternion algebra $\mathfrak{A} = (\mathfrak{Z}, S, \gamma)$ such that

$$(51) \quad \mathfrak{Z}_0 = \mathfrak{F}[i_0], \qquad i_0^2 = \beta_0 i_0 + \alpha_0 e, \qquad j_0^2 = \gamma_0 e \neq 0, \qquad j_0 i_0 + i_0 j_0 = j_0$$

for α_0, β_0, γ_0 in $\mathfrak{F}$. Applying (40) for an arbitrary generation we have

$$(52) \qquad j_0 x + x j_0 = T_{01}(x)e + T_{02}(x)j_0,$$

where $T_{01}(x)$ and $T_{02}(x)$ are in $\mathfrak{F}$. It follows that

$$\mathfrak{F}[j_0] \neq \mathfrak{Z}.$$

For otherwise $j_0 = \mu e + \lambda i$ where $\lambda \neq 0$ and μ are in $\mathfrak{F}$, $j_0 j + j j_0 = 2\mu j + \lambda j$ is not in $\mathfrak{Z}$, a contradiction.

If $\mathfrak{A}$ is nonsplit its only quadratic subalgebras are $\mathfrak{Z}$ and $\mathfrak{F}[j]$ and if $\mathfrak{F}[j_0] \neq \mathfrak{Z}$ then $\mathfrak{F}[j_0] = \mathfrak{F}[j]$, $\mathfrak{Z}_0 = \mathfrak{Z}$. When the characteristic of $\mathfrak{F}$ is not two then $j_0 = \lambda j$ for $\lambda \neq 0$ in $\mathfrak{F}$ and the result of this generation $\mathfrak{A} = \mathfrak{Z} + (\lambda j)\mathfrak{Z}$ is given by $(\alpha, \gamma) \cong (\alpha, \lambda^2\gamma)$. If the characteristic of $\mathfrak{F}$ is two then $\mathfrak{F}[j_0] = \mathfrak{F}[j]$ for any $j_0 = \mu e + \lambda j$ and $\lambda \neq 0$, μ in $\mathfrak{F}$. Then $i_0 = \alpha_1 e + \alpha_2 i$ where α_1 and α_2 are in $\mathfrak{F}$, $j_0 i_0 + i_0 j_0 = (\mu e i_0 + i_0 \mu e) + (\lambda j)(\alpha_1 e) + (\alpha_1 e)\lambda j + \alpha_2\lambda(ji + ij) = \alpha_2\lambda j = \mu e + \lambda j$ if and only if $\mu = 0$, $j_0 = \lambda j$, $\alpha_2 = 1$. The result of this generation is given by (26).

Assume next that $\mathfrak{A}$ is a split algebra over $\mathfrak{F}$ of characteristic two. By the argument above the only generation with $\mathfrak{F}[j_0] = \mathfrak{F}[j]$ and $\mathfrak{Z}_0 = \mathfrak{Z}$ is that resulting in (26), (27). By (53) and Theorem 7 we necessarily have $\mathfrak{F}[j_0] = \mathfrak{F}[j]$, $j_0 = \mu e + \lambda j$ for $\lambda \neq 0$ and μ in $\mathfrak{F}$. When $\mathfrak{Z}_0 \neq \mathfrak{Z}$ we have $\mathfrak{Z}_0 = \mathfrak{F}[x]$ with x given by (50). Then $i_0 = \alpha_1 x + \alpha_2 e$, $i_0 j_0 + j_0 i_0 = \lambda\alpha_1(xj + jx) = \lambda\alpha_1(ji + ij + i + e - i) = \lambda\alpha_1(e + j)$. Then $i_0 j_0 + j_0 i_0 = j_0$ if and only if $\lambda\alpha_1(e + j) = \mu e + \lambda j$, $\alpha_1 = 1$, $\lambda = \mu$, $j_0 = \lambda(e + j)$. But then $j_0^2 = 0$ contrary to our hypothesis that $\gamma_0 \neq 0$. Indeed $j_0 x = 0$ and $\mathfrak{A} \neq \mathfrak{F}[x] + j_0\mathfrak{F}[x]$. We have proved that there are no other generations other than that given by $\mathfrak{A} = \mathfrak{Z} + (\lambda j)\mathfrak{Z}$.

Our final case is that given by $\mathfrak{A} = (\alpha, 1)$ over $\mathfrak{F}$ of characteristic not two. Since $\mathfrak{F}[j_0] \neq \mathfrak{Z}$ we may use Theorem 6 to obtain $\mathfrak{F}[j_0] = \mathfrak{F}[x]$ where $x = \eta v + \epsilon\eta j v + \delta j$ such that $\delta^2 = \epsilon^2 = 1$. Moreover we may assume that $j_0^2 = e$ and so $j_0 = x$. But $jv + vj = 0$ and $(jv)v + v(jv) = 0$, $xv + vx = \eta v^2 = \eta(e - 2i)^2$ $\eta(e + 4\alpha e - 4i)$. This is not in $\mathfrak{F}[x]$ unless $\eta = 0$, $x = \pm j$. We have proved our main

THEOREM 8. *The only generations* (51) *of a quasiquaternion algebra* $(\mathfrak{Z}, S, \gamma)$ *are those in which* $\mathfrak{Z}_0 = \mathfrak{Z}$, $j_0 = \lambda j$ *for* $\lambda \neq 0$ *in* $\mathfrak{F}$.

10. Automorphisms and isomorphism

An isomorphism T of a quasiquaternion algebra $\mathfrak{A}_0 = (\mathfrak{Z}_0, S_0, \gamma_0)$ on a quasiquaternion algebra $\mathfrak{A} = (\mathfrak{Z}, S, \gamma)$ carries $\mathfrak{Z}_0$ into a quadratic subalgebra $\mathfrak{Z}_0 T$ of $\mathfrak{A}$, and j_0 into a quantity $j_0 T$ of $\mathfrak{A}$, such that $\mathfrak{A} = \mathfrak{Z}_0 T + (j_0 T)(\mathfrak{Z}_0 T)$ and $(j_0^2 T)^2 = \gamma_0 e$, $\mathfrak{Z}_0 T = \mathfrak{F}[i_0 T]$, $(i_0 T) j_0 T + (j_0 T) i_0 = j_0 T$. By Theorem 8 $\mathfrak{Z}_0 T = \mathfrak{Z}$, T induces an isomorphism of $\mathfrak{Z}$ and $\mathfrak{Z}_0$. Also $j_0 = \lambda j$ for $\lambda \neq 0$ in $\mathfrak{F}$. By the proof of Theorem 8 we have

THEOREM 9. *Two quasiquaternion algebras* (α, γ) *and* (α_0, γ_0) *over* $\mathfrak{F}$ *of characteristic not two are isomorphic if and only if* $\alpha_0 = \alpha$, $\gamma_0 = \lambda^2\gamma$ *for* $\lambda \neq 0$ *in* F. *Two algebras* (α, β, γ) *and* $(\alpha_0, \beta_0, \gamma_0)$ *over* $\mathfrak{F}$ *of characteristic two are isomorphic if and only if* $\beta_0 = \beta$, $\alpha_0 = \alpha + \mu^2 + \mu\beta$, $\gamma_0 = \lambda^2\gamma$ *for* $\lambda \neq 0$ *and* μ *in* $\mathfrak{F}$.

An automorphism of (α, γ) over $\mathfrak{F}$ of characteristic not two is an isomorphism which replaces i by itself and j by λ, such that $\lambda^2\gamma = \gamma$, $\lambda^2 = 1$. Then $\lambda = \pm 1$ and we have proved

THEOREM 10. *The automorphism group of* (α, γ) *over* $\mathfrak{F}$ *of characteristic not two is the group* (I, A) *of order two defined by* (20).

We also see that an automorphism of $(\alpha, 0, \gamma)$ over $\mathfrak{F}$ of characteristic two replaces i by $i_0 = i + \mu e$, j by λj such that $\lambda^2\gamma = \gamma$, $\lambda^2 = 1$, $\lambda = 1$, $(i + \mu e)^2 = \alpha + \mu^2 e = \alpha$, $\mu = 0$, $i_0 = i$. If $\mathfrak{A} = (\alpha, \beta, \gamma)$ an automorphism replaces i by $i_0 = i + \mu e$ where $i_0^2 = \beta i_0 + \mu e$, $j_0 = \lambda$, such that $\lambda^2\gamma = \gamma$, $\lambda = 1$. Then $i_0 = \beta e - i$ or i. We have proved

THEOREM 11. *Let* $\mathfrak{F}$ *be a field of characteristic two. Then the automorphism group of* $(\alpha, 0, \gamma)$ *is the identity group, the automorphism group of* (α, β, γ) *for every* $\beta \neq 0$ *is the group* I, B *of order two given by* B *defined in* (24).

It is a curious fact that while the antiisomorphs of quasiquaternion algebras are algebras of a similar type they are not quasiquaternion algebras. For let $\mathfrak{A}_0 = (\mathfrak{Z}_0, S_0, \gamma_0)$ be antiisomorphic to $\mathfrak{A} = (\mathfrak{Z}, S, \gamma)$ under a mapping $x_0 \to x_0 T$ on $\mathfrak{A}_0$ to $\mathfrak{A}$ such that $(x_0 c_0)T = (c_0 T)(x_0 T)$. Then $\mathfrak{Z}_0 T = \mathfrak{Z}_1$ is a quadratic subalgebra of $\mathfrak{A}$ isomorphic to $\mathfrak{Z}_0$, $\mathfrak{A} = \mathfrak{Z}_1 + Z_1 j_1$ where $j_1 = j_0 T$, $\mathfrak{Z}_1 = \mathfrak{F}[i_1]$, $j_1 i_1 = (e - i_1) j_1$. Then $j_1 i_1 + i_1 j_1 = j_1$, $i_1 j_1 = j_1(e - i_1)$. It follows from Theorem 6 that $\mathfrak{Z}_1 = \mathfrak{Z}$, $z_0 \to z_0 T$ is an isomorphism of $\mathfrak{Z}_0$ and $\mathfrak{Z}$, $j_1 = \lambda j$ for $\lambda \neq 0$ in $\mathfrak{F}$. But $(j_0 z_0) y_0 = j_0(z_0 y_0)$ for every y_0 and z_0 of $\mathfrak{Z}_0$, $[(j_0 z_0) y_0]T = y_0 T[(z_0 T)(y_0 T)] = y_1(z_1 j_1) = [(z_0 y_0)T](j_0 T) = (z_1 y_1) j_1$. However $(z_1 y_1) j_1 = \lambda j (z_1 y_1)'$, $y_1(z_1 j_1) = \lambda j (z_1 y_1)'$, $y_1(z_1 j_1) = \lambda j (z_1' y_1')$ and $(z_1 y_1)' = z_1' y_1'$, contrary to our assumption that S is not an automorphism of $\mathfrak{Z}$.

11. Division algebras

The algebras $\mathfrak{Z}$ and $\mathfrak{F}[j]$ are commutative and associative algebras and if either contains a divisor of zero so does $\mathfrak{A}$. Hence $\mathfrak{A}$ is a division algebra only if $\mathfrak{Z}$ and $\mathfrak{F}[j]$ are fields. Conversely let $\mathfrak{Z}$ and $\mathfrak{F}[j]$ be fields, $j^2 = \gamma e$. Then $\gamma \neq \delta^2$ for any of $\mathfrak{F}$, $N(x) = yy' - zz' = 0$ only if (19) holds. Then y and z are in

$\mathfrak{F}[e]$, $y' = y$, $z' = z$, $y^2 - \gamma z^2 = 0$, that is $x = 0$. This proves that $\mathfrak{A}$ is a division algebra. We have proved

THEOREM 12. *A quasiquaternion algebra is a division algebra if and only if its two quadratic subalgebras are fields.*

For the algebras of Theorem 10 and the algebras $(\alpha, 0, \gamma) = (\alpha, \gamma)$ this theorem states that (α, γ) is a division algebra if and only if neither α nor γ is the square of a quantity of $\mathfrak{F}$. For those of Theorem 11 with $\beta \neq 0$ it states that (α, β, γ) is a division algebra if and only if γ is not the square of any quantity of $\mathfrak{F}$ and $\alpha \neq \mu^2 - \beta\mu$ for any μ of $\mathfrak{F}$.

A scalar extension field $\mathfrak{K}$ of $\mathfrak{F}$ will be called an *intersecting* field for a division algebra $\mathfrak{A}$ over $\mathfrak{F}$ if $\mathfrak{A}_{\mathfrak{K}}$ is not a division algebra. From Theorem 12 we have

THEOREM 13. *A scalar extension $\mathfrak{K}$ of $\mathfrak{F}$ is an intersecting field of a quasiquaternion division algebra $\mathfrak{A}$ over $\mathfrak{F}$ if and only if $\mathfrak{K}$ contains a subfield over $\mathfrak{F}$ isomorphic to one of the two quadratic subalgebras of $\mathfrak{A}$.*

If $\mathfrak{F}$ is a finite field of characteristic two every quantity of $\mathfrak{F}$ is a square. Thus there are no quasiquaternion division algebras, and indeed no nonsplit algebras, over such a field. A finite field $\mathfrak{F}$ of characteristic $p \neq 2$ contains $q = p^\nu$ elements and there is only one quadratic extension of $\mathfrak{F}$. It is the field of q^2 elements and $\mathfrak{F}$ contains a quantity γ which is not a square in $\mathfrak{F}$ such that if $j^2 = \gamma e$ then $\mathfrak{F}[j]$ is a field. Necessarily $\alpha = \tau^2\gamma$, $i^2 = \alpha e$ and there are precisely $\frac{1}{2}(q - 1)$ distinct nonzero values of τ^2. By Theorem 9 two algebras defined for distinct values of α are nonisomorphic and we have proved

THEOREM 14. *Let $\mathfrak{F}$ be a finite field of q elements where q is odd. Then there exist $\frac{1}{2}(q - 1)$ quasiquaternion division algebras over $\mathfrak{F}$ such that no two of these algebras are isomorphic.*

12. Isotopy of division algebras

Let x be the general quantity of any algebra $\mathfrak{A}$, L_x be the left multiplication of x as a quantity of $\mathfrak{A}$. Then the left multiplications $L_x^{(0)}$ of any isotope $\mathfrak{A}_0$ with a unity quantity of $\mathfrak{A}$ are given[6] by

$$L_{xQ}^{(0)} = PL_c^{-1}L_xP^{-1},$$

where P and Q are fixed nonsingular linear transformations on $\mathfrak{A}$, c is a fixed left nonsingular quantity of $\mathfrak{A}$. If $\mathfrak{A}$ is a quasiquaternion algebra and we take $x = jc$ we use (37) to write $L_c^{-1}L_x = L_c^{-1}L_cL_j = L_j$. Then $L_{(jc)}Q^{(0)} = PL_jP^{-1}$ and we have

LEMMA 5. *Every isotope with a unity quantity of a quasiquaternion algebra $\mathfrak{A} = (\mathfrak{Z}, S, \gamma)$ contains a quantity j_0 such that $(L_{j_0}^{(0)})^2 = \gamma I$.*

We also have

LEMMA 6. *Let $\mathfrak{A} = (\mathfrak{Z}, S, \gamma)$ be a nonsplit quasiquaternion algebra. Then the only quantities x of $\mathfrak{A}$ such that L_x^2 is a scalar transformation are the quantities x in $\mathfrak{F}[j]$ with this property.*

[6] For the definition of an isotope and the fundamental properties see my *Nonassociative algebras* I. *Fundamental concepts and isotopy*, these Annals vol. 43 (1942), pp. 685–707.

For if L_x^2 is a scalar δI then $eL_x^2 = x^2 = \delta e$. If x is not in $\mathfrak{F}[j]$ then $\mathfrak{F}[x]$ is a quadratic subalgebra of $\mathfrak{A}$ and, by Theorem 1, $\mathfrak{F}[x] = \mathfrak{Z}$. But if x is in $\mathfrak{Z}$ then

$$L_x = \begin{pmatrix} x & 0 \\ 0 & x' \end{pmatrix}, \quad L_x^2 = \begin{pmatrix} x^2 & 0 \\ 0 & (x')^2 \end{pmatrix} = \delta I$$

only if $x^2 = x'^2 = \delta e$. However $x^2 = x'^2$ if and only if $x' = \pm x$. In the former case $x' = x$ is in $\mathfrak{F}[e]$ and hence in $\mathfrak{F}[j]$. In the latter case $\mathfrak{F}$ has characteristic not two and $x' = \eta v = \eta(1 - 2i)$, $x^2 = \eta^2(1 - 4i + 4\alpha) \neq \delta e$.

As a consequence of the proofs of our lemmas we have

THEOREM 15. *Every quasiquaternion isotope $\mathfrak{A}_0 = (\mathfrak{Z}_0, S_0, \gamma_0) = \mathfrak{Z}_0 + j_0\mathfrak{Z}_0$ of a nonsplit quasiquaternion algebra $\mathfrak{A} = (\mathfrak{Z}, S, \gamma) = \mathfrak{Z} + j\mathfrak{Z}$ is nonsplit and $\mathfrak{F}[j_0] = \mathfrak{F}[j_1]$ such that $(L_{j_1}^{(0)})^2 = \gamma I$.*

For by the proof of Lemma 5 we have $(L_{j_1}^{(0)})^2 = \gamma I$ where

$$j_1 = (jc)Q.$$

Then $\mathfrak{F}[j_1]$ is a field. By Theorems 6 and 7 the only subfield of a split quasiquaternion algebra is $\mathfrak{Z}$ and, by the proof of Lemma 6, no x in $\mathfrak{Z}_0$ and not in $\mathfrak{F}[e]$ has the property that $(L_x^{(0)})^2 = \gamma I$. Hence $\mathfrak{A}_0$ cannot be a split algebra, $\mathfrak{F}[j_1]$ is a quadratic subfield of the quadratic field $\mathfrak{F}[j_0]$, $\mathfrak{F}[j_1] = \mathfrak{F}[j_0]$.

We now observe that if the characteristic of $\mathfrak{F}$ is not two then necessarily

$$j_1 = \lambda j_0$$

where $\lambda \neq 0$ is in $\mathfrak{F}$. By (39) we have $L_{x_0}^{(0)}L_{j_1}^{(0)} + L_{j_1}^{(0)}L_{x_0}^{(0)} = \eta_1 I + \eta_2 L_{j_1}^{(0)}$ for every x_0 in $\mathfrak{A}_0$ where η_1 and η_2 are in $\mathfrak{F}$. If $\mathfrak{F}$ has characteristic two this result also holds since $j_1 = \lambda_0 e_0 + \lambda j_0$ and $L_{x_0}^{(0)}L_{j_1}^{(0)} + L_{j_1}^{(0)}L_{x_0}^{(0)} = \lambda(L_{x_0}^{(0)}L_{j_0}^{(0)} + L_{j_0}^{(0)}L_{x_0}^{(0)})$. But then, multiplying by P^{-1} on the left and P on the right, we have

$$L_j(L_c^{-1}L_x) + (L_c^{-1}L_x)L_j = (\eta_1 I + \eta_2 L_j)$$

for η_1 and η_2 in $\mathfrak{F}$. Take $x = c$ to obtain the result used in (43) to imply Theorem 3. We have proved

THEOREM 16. *Let $\mathfrak{A}$ and $\mathfrak{A}_0$ be isotopic nonsplit quasiquaternion algebras. Then there exist nonsingular transformations P and Q on $\mathfrak{A}$ such that*

$$L_{xQ}^{(0)} = PL_xP^{-1}$$

for every x of $\mathfrak{A}$.

We next prove

THEOREM 17. *If $\mathfrak{A} = (\mathfrak{Z}, S, \gamma)$ and $\mathfrak{A}_0 = (\mathfrak{Z}_0, S_0, \gamma_0)$ are isotopic quasiquaternion division algebras the fields $\mathfrak{Z}$ and $\mathfrak{Z}_0$ are isomorphic.*

For we may construct scalar extensions $\mathfrak{K}$ and $\mathfrak{K}_0$ isomorphic over $\mathfrak{F}$ respectively to $\mathfrak{Z}$ and $\mathfrak{Z}_0$. The algebra $\mathfrak{A}_\mathfrak{K}$ contains a quantity x such that L_x is singular, PL_xP^{-1} is singular, $(\mathfrak{A}_0)_\mathfrak{K}$ is not a division algebra. Similarly $\mathfrak{A}_{\mathfrak{K}_0}$ is not a division algebra. By Theorem 13 the field $\mathfrak{K}$ is isomorphic either to $\mathfrak{Z}_0$ or to

$\mathfrak{F}[j_0]$, the field $\mathfrak{K}_0$ is isomorphic either to $\mathfrak{Z}$ or to $\mathfrak{F}[j]$. Since $\mathfrak{F}[j]$ and $\mathfrak{F}[j_0]$ are isomorphic these isomorphisms imply in all cases that $\mathfrak{Z}$ is isomorphic to $\mathfrak{Z}_0$.

If $\mathfrak{F}$ has characteristic two Theorem 15 implies only that $j_1 = \lambda_0 e_0 + \lambda j_0$ for λ_0 and $\lambda \neq 0$ in $\mathfrak{F}$. However if the characteristic of $\mathfrak{F}$ is not two the property $j_1^2 = \gamma e_0$ implies that $j_1 = \lambda j_0$. Since we shall not complete the study of isotopy of quasiquaternion algebras here we shall restrict our attention to the case of division algebras over $\mathfrak{F}$ of characteristic not two. Then $\mathfrak{A} = (\alpha, \gamma)$, $\mathfrak{A}_0 = (\alpha_0, \gamma_0)$ such that $j_1^2 = (\lambda j_0)^2 = \lambda^2\gamma_0 e = \gamma e$. Then by (25) $\mathfrak{A}_0 = (\alpha_0, \gamma)$ and we may assume that $\gamma_0 = \gamma$.

We may now use the basis e, i, j, ji of $\mathfrak{A}$ over $\mathfrak{F}$ as a basis of $\mathfrak{A}_0$. Then multiplication in $\mathfrak{A}_0$ is defined by $x \cdot c = cL_x^{(0)}$ such that $j \cdot i = ji$, $i \cdot i = \alpha_0 e$, $j \cdot j = \gamma e$, $j \cdot i + i \cdot j = j$. It follows that

$$L_j^{(0)} = L_j, \qquad L_i^{(0)} = \begin{pmatrix} i_0 & 0 \\ 0 & e - i_0 \end{pmatrix}.$$

By Theorem 17 we have $\alpha_0 = \mu^2\alpha$ and we may take

$$i_0 = \mu i$$

for $\mu \neq 0$ in $\mathfrak{F}$.

We now observe that $\lambda^2\gamma = \gamma$ so that $\lambda = \pm 1$, $j_1 = \pm j$, $PL_jP^{-1} = \epsilon L_j$, $\epsilon = \pm 1$. Write

$$P = \begin{pmatrix} P_1 & P_2 \\ P_3 & P_4 \end{pmatrix}, \qquad PL_j = \begin{pmatrix} \gamma P_2 & P_1 \\ \gamma P_4 & P_3 \end{pmatrix} \qquad L_jP = \begin{pmatrix} P_3 & P_4 \\ \gamma P_1 & \gamma P_2 \end{pmatrix}.$$

It follows that $P_4 = \epsilon P_1$, $P_3 = \epsilon\gamma P_2$ and thus that

$$P = \begin{pmatrix} P_1 & P_2 \\ \epsilon\gamma P_2 & \epsilon P_1 \end{pmatrix}.$$

Suppose now that $L_i^{(0)} = PL_cP^{-1}$ where $c = a + jb$. By (39) applied in $\mathfrak{A}_0$ we have $L_i^{(0)}L_j^{(0)} + L_j^{(0)}L_i^{(0)} = L_j^{(0)}$ and thus $L_cL_j + L_jL_c = L_j$. It follows from (40) that $T(b) = 0$, $T(a) = 1$. By Lemma 1

$$b = \eta v, \qquad a = i + \xi v, \qquad L_c = \begin{pmatrix} a & b \\ -\gamma b & e - a \end{pmatrix}.$$

We now compute PL_c and $L_i^{(0)}P$ to obtain the relations

$$P_1a - \gamma P_2b = \mu i P_1, \qquad P_1b - P_2(e - a) = \mu i P_2 \tag{53}$$

as well as two other, but equivalent, relations. Then $\gamma P_2bP_1a - \mu iP_1$, $\gamma P_1b^2 + (P_1a - \mu iP_1)(e - a) = \mu i(P_1a - \mu iP_1)$. Then $P_1[\gamma b^2 + a(e - a)] = \mu iP_1 - \mu^2\alpha P_1$. It follows that

(54) $$\mu i P_1 = P_1 y, \qquad y = \gamma b^2 + a(e - a) + \mu^2 \alpha e,$$

so that y is in $\mathfrak{Z}$. Also by (53)

(55) $$P_1(a - y) = P_2 b.$$

If $b = 0$ then $P_1 a = \mu i P_1$ and $P_1 a^2 = (\mu i)^2 P_1 = P_1(\alpha_0 e)$, $P_1(a^2 - \alpha_0 e) = 0$. If $P_1 \neq 0$ then our hypothesis that $\mathfrak{Z}$ is a field implies that $a^2 = \alpha_0 e$, $a = \pm \mu i$. However $a' = \pm\mu(e - i)$, $a + a' = \pm\mu e = e$, $u = \pm 1$, $\mathfrak{A}_0 = (\alpha, \gamma)$ is isomorphic to $\mathfrak{A}$. If $P_1 = 0$ then $P_2 \neq 0$, $P_2 a' = (-\mu i)P_2$, $P_2(a')^2 = (-\mu i)^2 P_2 = P_2(\alpha_0 e)$, $a'^2 = \alpha_0 e$, $a' = \pm \mu i$, $(a')' = \pm \mu(e - i) = e - a'$, $a' + a = \pm \mu e = e$, $\mu = \pm 1$ and again $\mathfrak{A}$ and $\mathfrak{A}_0$ are isomorphic.

Hence let $b \neq 0$. Then

(56) $$P_2 = P_1 z, \qquad z = (a - y)(\gamma b)^{-1},$$

and

(57) $$P = \begin{pmatrix} P_1 & 0 \\ 0 & P_1 \end{pmatrix} \begin{pmatrix} e & z \\ e\gamma z & \epsilon e \end{pmatrix}$$

is nonsingular if and only if P_1 is nonsingular and $e - \gamma z^2 \neq 0$. However $e - \gamma z^2 = e - \gamma(a - y)^2(\gamma b^{-1})^2 \neq 0$ if and only if

(58) $$(a - y)^2 \neq \gamma b^2.$$

Since P_1 is nonsingular the quantity $y = P_1^{-1}\mu i P_1$ has the property that $y^2 = \mu^2 \alpha e$. But y is in $\mathfrak{Z}$ and thus $y = \pm \mu i$. Since only μ^2 is determined by $\mathfrak{A}_0$ we may assume the notation μ selected so that $y = \mu i$. We compute $y = \gamma\eta^2(1 + 4\alpha - 4i)e + i + \xi(e - 2i) - [\alpha e + 2i(1 - 2i)\xi + \xi^2(1 - 4i + 4\alpha)e] + \alpha\mu^2 e$ and see that $y = \mu i$ if and only if

(59) $$\mu = 4(\xi^2 - \xi - \gamma\eta^2) + 1, \qquad (1 + 4\alpha)(\xi - \xi^2 + \gamma\eta^2) = \alpha(1 - \mu^2).$$

Then $1 - \mu = 4(\xi - \xi^2 + \gamma\eta^2)$, $(1 + 4\alpha)(1 - \mu) = 4\alpha(1 - \mu^2)$, $(1 - \mu)(1 - 4\alpha\mu) = 0$ and $\mu = 1$ or $(4\alpha)^{-1}$. Note that the first equation of (59) is equivalent to

(60) $$(1 - 2\xi)^2 - 4\gamma\eta^2 = \mu.$$

If $\mu = (4\alpha)^{-1}$ in (60) there exist quantities $\zeta_1 = (1 - 2\xi)2\alpha$, $\zeta_2 = 4\eta\alpha$ such that $\alpha = \zeta_1^2 - \gamma\zeta_2^2$, $\gamma\zeta_2^2 = \zeta_1^2 - \alpha$, γ is the norm of a quantity of the field $\mathfrak{Z}$. We have proved

THEOREM 18. *Let $\mathfrak{A} = (\alpha, \gamma) = (\mathfrak{Z}, S, \gamma)$ be a quasiquaternion division algebra over $\mathfrak{F}$ of characteristic not two such that γ is not the norm of any quantity of $\mathfrak{Z}$. Then a quasiquaternion algebra is isotopic to $\mathfrak{A}$ if and only if it is isomorphic to $\mathfrak{A}$.*

Suppose next that γ is a norm and so $\alpha = \zeta_1^2 - \zeta_2^2\gamma$ for ζ_1 and ζ_2 in $\mathfrak{F}$. Then ξ and η may be determined so that $\zeta_1 = (1 - 2\xi)2\alpha$, $\zeta_2 = 4\eta\alpha$ and (60) is satisfied for $\mu = (4\alpha)^{-1}$. Then (59) is satisfied and there exist quantities a and b in $\mathfrak{Z}$ such that $\mu_i = \gamma b^2 + a - a^2 + \mu^2\alpha e$. If $(a - \mu i)^2 = \gamma b^2$ then $a^2 - 2\mu i a + \mu 2\alpha e$

$= \mu i + a^2 - a - u^2\alpha e, 2\mu^2\alpha e - \mu i = (2\mu i - e)a, (2\mu i - e)\mu i = (2\mu i - e)a,$ $a = \mu i$. But $a = i + \xi(e - 2i)$ and $a = \mu i$ if and only if $\xi = 0$, $\mu = 1$ a contradiction. Hence the transformation

$$P = \begin{pmatrix} e & z \\ \gamma z, & e \end{pmatrix}$$

defined by $z = (a - \mu i)(\gamma b)^{-1}$ is nonsingular and

$$PL_cP^{-1} = \begin{pmatrix} i & 0 \\ 0 & e - ui \end{pmatrix} = L_i^{(0)}, \qquad PL_jP^{-1} = L_j.$$

Then $PL(\mathfrak{A})P^{-1}$ contains PL_xP^{-1} for every $x = \alpha_1 e + \alpha_{2c}$ and thus contains $L^{(0)}_{\alpha_1 e + \alpha_2 i}$. Since $L_{jc} = L_cL_j$ we have

$$PL_{jc}P^{-1} = PL_cL_jP^{-1} = \begin{pmatrix} i & 0 \\ 0 & e - \mu i \end{pmatrix} = \begin{pmatrix} 0 & e \\ \gamma e & 0 \end{pmatrix} = \begin{pmatrix} 0 & \mu i \\ \gamma(e - \mu i) & 0 \end{pmatrix}$$

in $PL(\mathfrak{A})P^{-1}$, $PL(\mathfrak{A})P^{-1} = L(\mathfrak{A}_0)$. We have proved

THEOREM 19. *Let $\mathfrak{A} = (\alpha, \gamma) = (\mathfrak{Z}, S, \gamma)$ be a quasiquaternion division algebra over $\mathfrak{F}$ of characteristic not two and γ be the norm of a quantity of $\mathfrak{Z}$. Then a quasiquaternion algebra is isotopic to $\mathfrak{A}$ if and only if it is either isomorphic to $\mathfrak{A}$ or to (α_0, γ) where $\alpha_0 = (16\alpha)^{-1}$.*

We shall conclude our study with a discussion of the real case. By Theorem 12 every quasiquaternion division algebra over the field of all real numbers is an algebra $\mathfrak{A} = (-\mu^2, -1)$, where μ is any *positive* real number. By Theorem 18 *two such algebras defined for distinct values of μ* are not merely nonisomorphic but *are actually nonisotopic.*

THE UNIVERSITY OF CHICAGO

ANNALS OF MATHEMATICS
Vol. 45, No. 4, October, 1944

ON THE DECOMPOSITION OF MODULAR TENSORS (II)

BY R. M. THRALL

(Received October 14, 1943)

Introduction

The present paper is the second part of [8] (brackets refer to bibliography) referred to below as Part I. We number sections, formulae, theorems, etc. consecutively from those in Part I, and use the same notation. For instance, $\mathfrak{f}$ is a field of characteristic p with q ($\leqq \infty$) elements; $\mathfrak{B}_1$ is a $\mathfrak{f}$-vector space of dimension n; $\mathfrak{G} = \mathfrak{G}(n, \mathfrak{f})$ is the full linear (modular) group of all $\mathfrak{f}$-automorphisms of $\mathfrak{B}_1$; $\mathfrak{B}_m$ is the space of all tensors of rank m over $\mathfrak{B}_1$; Π_m is the Kronecker m^{th} power representation of $\mathfrak{G}$; $\mathfrak{A}_m$ is the enveloping algebra of Π_m. Superscripts zero indicate the analogous quantities defined over a field $\mathfrak{f}^0$ of characteristic zero.

The decomposition of non-modular tensors, or equivalently the determination of the reduced form of Π_m^0 is now a classical part of algebra. The objective of the present sequence of papers is to obtain a similar theory for modular tensors, or equivalently to study the reduced form of Π_m. Any representation of $\mathfrak{G}$ whose space is a subspace or factor space of $\mathfrak{B}_m$, or a direct sum of such spaces we call a *tensor representation* of $\mathfrak{G}$. If q (and therefore $\mathfrak{G}$) is finite we denote the group ring by Γ. One of the main results of the present paper is that there exists a faithful tensor representation of Γ (Th. X §14). From this it follows from an unpublished theorem of Nesbitt that every representation of $\mathfrak{G}$ is equivalent to a tensor representation, but we do not establish or apply this last result below.

An important feature of the study of modular representations of groups has been the use of "induced representations". One starts with a finite group and a non-modular representation of it, and after suitable number theoretic preparations take residue classes and obtain a modular representation. A generalization of this process would be to induce both the group and the representation. We leave for future investigation the determination of the general theory of such a process, and content ourselves here with the application of the idea to obtain from each irreducible representation of the non-modular full linear group a representation of the modular full linear group. This is done in §9 below. In §10 a character theory is developed for tensor representations of $\mathfrak{G}$; this is applied in §11 to prove that every irreducible representation of $\mathfrak{G}$ is equivalent to a tensor representation,[1] and in §§12, 13 to obtain specific values of the irreducible and indecomposable modular characters for the representations Π_m having $m < 2p$.

In Part I the reduced form of Π_m for $m < 2p$ was determined subject to the assumption $q \geqq m$. In §§15–20 the situation for $m < 2p$ is cleaned up by re-

[1] Since submitting this paper the author has learned that this result was also obtained by Carson Mark in his Toronto thesis (1938, as yet unpublished).

moving this assumption and finding what further reductions can be made when $q = p$. A main tool in this investigation is the character theory of §13. The most troublesome case is for $m = 2p - 1, n = 2$. The difficulty (and also special interest) of this case is due to the influence of zero divisors in the ring of functions over a finite field. This case is the only one for $m < 2p$ where these zero divisors need be considered; but for larger values of m they will enter much of the time.

§9. Induced tensor representations

In the present section we show how to construct a modular tensor representation from each ordinary irreducible tensor representation. (We use the term "tensor representation" to refer to a representation of the full linear group whose space is a subspace or factor space of the space of all tensors of some rank, or for a direct sum of such representations). We follow the notation of Part I for modular representations; i.e. $\mathfrak{B}_1$ is a space (of dimension n) over a field $\mathfrak{k}$ of characteristic p. $\mathfrak{B}_m$ (the space of all tensors of rank m) is representation space for the Kronecker m^{th} power representation $\Pi_m: A \to \Pi_m(A) = A \times \cdots \times A$ (m factors) of the full linear group $\mathfrak{G}(= \mathfrak{G}(n, \mathfrak{k}))$ of all non-singular linear transformations A of $\mathfrak{B}_1$ onto itself. Let $\mathfrak{k}^0$ be a field of characteristic zero. We shall use superscripts zero to indicate analogous groups, spaces, and representations over $\mathfrak{k}^0$. Thus $\mathfrak{G}^0$ is the full linear group on a $\mathfrak{k}^0$-vector space $\mathfrak{B}_1^0$ (of dimension n), etc. In what follows we shall frequently choose for $\mathfrak{k}^0$ a field such that $\mathfrak{k}$ is residue class ring of the ring of integers in $\mathfrak{k}^0$ with respect to some prime ideal. When this is the case we shall say that $\mathfrak{k}^0$ is a field which *induces* $\mathfrak{k}$.

An irreducible constituent of Π_m^0 is characterized [10. P. 124, Th. 4.4D] by a partition of m. We write $\mathfrak{H}^0(\lambda): A^0 \to H^0(\lambda \wr A^0)$ for the irreducible constituent of Π_m^0 associated with the partition $(\lambda): \lambda_1 + \lambda_2 + \cdots + \lambda_k = m, \lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_k > 0$.

Lemma VII. *With proper choice of basis vectors in the representation space, $H^0(\lambda)$, of $\mathfrak{H}^0(\lambda)$, we can find polynomials $h_{uv}^0(X_{ik}^0)$ in n^2 indeterminants X_{ik}^0 and with rational integer coefficients such that the matrix $A^0 = \| a_{ik}^0 \|$ is represented by the matrix $H^0(\lambda \wr A^0) = \| h_{uv}^0(a_{ik}^0) \|$.*

Proof: We consider the $\mathfrak{k}^0$-group ring $\mathfrak{R}^0(= \mathfrak{R}^0(m, \mathfrak{k}^0))$ of the symmetric group $\mathfrak{S}_m$. In Part I [8. P. 675] we showed how to interpret the elements, s, of $\mathfrak{S}_m$ as linear operators on the tensor space $\mathfrak{B}_m^0$, and we now define $a = \sum a(s)s$, ($a(s)$ in $\mathfrak{k}^0$) as an operator on $\mathfrak{B}_m^0$ by linearity; i.e. $aX^0 = \sum a(s)sX^0$ for all vectors X^0 of $\mathfrak{B}_m^0$ and all elements a of $\mathfrak{R}^0$.

We start with the classical theorem that $H^0(\lambda)$ has the form $e\mathfrak{B}_m^0$ where e is a primitive idempotent of $\mathfrak{R}^0$ [10. P. 129, Th. 4.4D]. We also use the fact that e (as an element of $\mathfrak{R}^0$) is commutative with each operator $\Pi_m^0(A^0)$.

We consider the natural basis for the tensor space $\mathfrak{B}_m^0$ (i.e. the one relative to which the effect of A^0 is given by the Kronecker m^{th} power of A^0). Let $x(i_1, \cdots, i_m)$ denote the particular basis tensor which has its $i_1, \cdots, i_m$ component unity and all other components zero, and set $ex(i_1, \cdots, i_m) = x'(i_1, \cdots, i_m)$. Then

$H^0(\lambda)$ is the set of all $\mathfrak{k}^0$-combinations of the $x'(i_1, \cdots, i_m)$. Since $\Pi_m^0(A^0)e = e\Pi_m^0(A^0)$ we have from

$$\Pi_m^0(A^0)x(i_1, \cdots, i_m) = \sum_{\nu_1, \cdots, \nu_m} a_{i_1\nu_1} \cdot \cdots \cdot a_{i_m\nu_m} x(\nu_1, \cdots, \nu_m)$$

that

$$\Pi_m^0(A^0)x'(i_1, \cdots, i_m) = \sum_{\nu_1, \cdots, \nu_m} a_{i_1\nu_1} \cdot \cdots \cdot a_{i_m\nu_m} x'(\nu_1, \cdots, \nu_m). \tag{52}$$

Let $\mathfrak{J}^0$ denote the ring of rational integers. The set of all $\mathfrak{J}^0$-combinations of the $x'(i_1, \cdots, i_m)$ is an $\mathfrak{J}^0$-space, W^0, contained in $H^0(\lambda)$. It is a classical property of $\mathfrak{J}^0$-spaces (lattices) that W^0 has an $\mathfrak{J}^0$-basis which is also a $\mathfrak{k}^0$-basis for $H^0(\lambda)$ [9. P. 142]. (By "basis" we mean a linearly independent set of vectors which span a space.) Now our lemma follows from formula (52) if we express the $x'(i_1, \cdots, i_m)$ and the vectors of the given $\mathfrak{J}^0$-basis in terms of each other.

Let $\mathfrak{J}_p$ denote the field of integers mod p, and let X_{ik} be indeterminants over $\mathfrak{J}_p$. Denote by $h_{uv}(X_{ik})$ the polynomial obtained from $h_{uv}^0(X_{ik}^0)$ by replacing each coefficient by its residue class mod p and X_{ik}^0 by X_{ik}. Then we have:

LEMMA VIII. *Let $A = \| a_{ik} \|$ be any element of (the full modular linear group) $\mathfrak{G}$, and set $H(\lambda \wr A) = \| h_{uv}(a_{ik}) \|$. Then the correspondence $\mathfrak{H}(\lambda)$: $A \rightarrow H(\lambda \wr A)$ is a representation of $\mathfrak{G}$, which is referred to as the representation induced by $\mathfrak{H}^0(\lambda)$.*

PROOF: Let $A^0 = \| a_{ik}^0 \|$ and $B^0 = \| b_{ik}^0 \|$ be matrices whose elements are indeterminants over $\mathfrak{J}^0$ and set $A^0B^0 = C^0$, i.e. $c_{ik}^0 = \sum a_{i\nu}^0 b_{\nu k}^0$. Since $\mathfrak{H}^0(\lambda)$ is a representation of $\mathfrak{G}^0$ we have

$$h_{uv}^0(c_{ik}^0) = \sum_\rho h_{u\rho}^0(a_{ik}^0) h_{\rho v}^0(b_{ik}^0). \tag{53}$$

Let a_{ik}, b_{ik} be indeterminants over the field $\mathfrak{J}_p$. The mapping $a_{ik}^0 \rightarrow a_{ik}$, $b_{ik}^0 \rightarrow b_{ik}$, integer $\rightarrow$ its residue class mod p, is a homomorphism of the ring $\mathfrak{J}^0[a_{ik}^0, b_{ik}^0]$ onto $\mathfrak{J}_p[a_{ik}, b_{ik}]$. Under this homomorphism (53) is replaced by

$$h_{uv}(c_{ik}) = \sum_\rho h_{u\rho}(a_{ik}) h_{\rho v}(b_{ik}) \tag{54}$$

where $c_{ik} = \sum a_{i\nu} b_{\nu k}$. If we identify $\mathfrak{J}_p$ with the prime field of $\mathfrak{k}$ and specialize the a_{ik} and b_{ik} into elements of $\mathfrak{k}$ (such specialization is a homomorphism) then (54) is seen to be the condition that $\mathfrak{H}(\lambda)$ should be a representation of $\mathfrak{G}$.

Observe that the representation $\mathfrak{H}(\lambda)$ just constructed depends upon the particular idempotent e and after that upon the particular basis chosen for the $\mathfrak{J}^0$-space W^0. We do not contend that two different primitive idempotents e (belonging to the same (λ)) need give rise to equivalent $\mathfrak{H}(\lambda)$, although it is evident that two different choices of bases for W^0 will give rise to equivalent $\mathfrak{H}(\lambda)$. What we can expect is that all the $\mathfrak{H}(\lambda)$ (for a given (λ)) will have the same irreducible constituents and this can be proved by arguments strictly analogous to those used by R. Brauer for finite groups [2. P. 954]. We omit the proof.

Although $\mathfrak{H}^0(\lambda)$ is a tensor representation it is not evident that $\mathfrak{H}(\lambda)$ is. The

root of the difficulty is that the idempotent $e = \sum e(s)s$ will not in general have any image in $\mathfrak{R}$ (because the $e(s)$ need not be integral at the prime spots dividing p). We do not investigate the general problem of determining those induced representations which are tensor representations; the following lemma is sufficient for our needs.

LEMMA IX. *For each partition* (λ) *there is at least one induced representation* $\mathfrak{H}(\lambda)$ *which is a tensor representation.*

PROOF: For any (λ) we can find a primitive idempotent $e = \sum e(s)s$ where each coefficient is one of the numbers $1/h$, 0, or $-1/h$, h a rational integer [10. P. 125, Th. 4.3F]. In the proof of Lemma VII replace $x'(i_1, \cdots, i_m)$ by $x''(i_1, \cdots, i_m) = hx'(i_1, \cdots, i_m)$. The $\mathfrak{J}^0$-combinations of the $x''(i_1, \cdots, i_m)$ constitute an $\mathfrak{J}^0$-subspace W_1^0 of W^0, and every vector of W_1^0 is in the $\mathfrak{J}^0$-space generated by the basis tensors $x(i_1, \cdots, i_m)$. $he = \sum he(s)s$ has coefficients 1, 0, or -1 and so has a non-zero image ρ in the modular group ring $\mathfrak{R}$. The homomorphism argument used to establish Lemma VIII can now be used to show that $\rho\mathfrak{B}_m$ is representation space for the induced representation $\mathfrak{H}(\lambda)$. We omit the details.

§10. Characters

Following Brauer and Nesbitt [1. P. 559] we use complex numbers as characters for modular representations. For the ensuing argument we temporarily restrict ourselves to a finite field $\mathfrak{k}$ so that $\mathfrak{G}$ will be a finite group. However, the results obtained are easily extended to infinite fields. Suppose that the order of $\mathfrak{G}$ is $g = p^a g'$ where $(g', p) = 1$. Let $\mathfrak{k}'$ be the splitting field over $\mathfrak{k}$ for the equation $x^{g'} = 1$, and take for $\mathfrak{k}^0$ a field which induces $\mathfrak{k}'$. Let ϵ be a primitive g'^{th} root of unity in $\mathfrak{k}'$ and let w be a complex number primitive g'^{th} root of unity. (We may suppose that w is in $\mathfrak{k}^0$ and that ϵ is the image in $\mathfrak{k}'$ of w.)

Let $\mathfrak{T}$: $A \to T(A)$ be any modular representation of $\mathfrak{G}$. The eigen-values of $T(A)$ will all be powers of ϵ and if A is p-regular (i.e. of order prime to p) $T(A)$ is similar to a diagonal matrix of the form

$$\left\| \begin{matrix} \epsilon^{\alpha_1} & & & \\ & \epsilon^{\alpha_2} & & \\ & & \cdot & \\ & & & \cdot \\ & & & & \cdot \end{matrix} \right\|$$

We call $\chi_{\mathfrak{T}}(A) = w^{\alpha_1} + w^{\alpha_2} + \cdots$ the *character* of the representation $\mathfrak{T}$. Brauer and Nesbitt have proved that a necessary and sufficient condition for two modular representations $\mathfrak{T}$ and $\mathfrak{T}'$ to have the same irreducible constituents is that $\chi_T(A) = \chi_{T'}(A)$ for all p-regular A in $\mathfrak{G}$ [1. P. 560].

Our program is to calculate the character $\chi_{(\lambda)}(A)$ of the representation $\mathfrak{H}(\lambda)$, and then use it to get information about the irreducible and indecomposable tensor characters of $\mathfrak{G}$. We may consider $\mathfrak{H}(\lambda)$ as being a representation of the group $\mathfrak{G}' = \mathfrak{G}(n, \mathfrak{k}')$. Any element A of $\mathfrak{G}'$ can be regarded as the residue class

image of some matrix A^0 in the group $\mathfrak{G}^0 = \mathfrak{G}(n, \mathfrak{k}^0)$. If two elements A_1, A_2 of $\mathfrak{G}'$ are similar the same is true of $T(A_1)$, $T(A_2)$ for any representation $\mathfrak{T}$ of $\mathfrak{G}'$ (for similarity in the full linear group is demonstrated by inner isomorphisms). Since every p-regular element of $\mathfrak{G}$ is similar to a diagonal element of $\mathfrak{G}'$ we can compute the character of $\mathfrak{H}(\lambda)$, considered as a representation of $\mathfrak{G}$, provided we know the value of $\chi_{(\lambda)}(A)$ for the similar diagonal $\mathfrak{k}'$-matrices A.

Suppose $A = \operatorname{diag}(a_1, a_2, \cdots)$ where $a_i = \epsilon^{\alpha_i}$. Then set $A^0 = \operatorname{diag}(a_1^0, a_2^0, \cdots)$ where $a_i^0 = w^{\alpha_i}$. The homomorphism of the ring of integers in $\mathfrak{k}^0$ onto $\mathfrak{k}'$ sends A^0 into A, by Lemma VII sends $H^0(\lambda\{A^0)$ into $H(\lambda\{A)$, and so sends the eigenvalues of $H^0(\lambda\{A^0)$ into the eigenvalues of $H(\lambda\{A)$. The eigenvalues of $H^0(\lambda\{A^0)$ are known to be power products of the a_i^0 and so those of $H(\lambda\{A)$ will be the same power products of the a_i. Therefore, the mapping $\epsilon^\alpha \to w^\alpha$ sends the eigenvalues of $H(\lambda\{A)$ back into those of $H^0(\lambda\{A^0)$ and so we have [3. P. 84; 10. P. 203].

$$\chi_{(\lambda)}(A) = \{\lambda(A^0)\} = \operatorname{tr} H^0(\lambda\{A^0). \tag{55}$$

In other words the character of $\mathfrak{H}(\lambda)$ (for $\mathfrak{G}$) is the well known Schur function of the complex number counterimages of the modular eigenvalues. We set $\chi_{(\lambda)}(A) = \{\lambda(A)\}$, and when the argument A is unimportant we drop it leaving just $\{\lambda\}$.

In Part I it was proved that for a given tensor representation (such as the $\mathfrak{H}(\lambda)$ of Lemma IX) the decomposition into irreducible and indecomposable constituents was independent of the field $\mathfrak{k}$, provided merely that $\mathfrak{k}$ was not too small [8. P. 675, Th. III]. Hence formula (55) can be used in determining the irreducible constituents even though the field $\mathfrak{k}$ is infinite.

§11. Irreducible representations

THEOREM VIII. *If $\mathfrak{k}$ is any finite field then every irreducible (modular) representation of the full linear group $\mathfrak{G} = \mathfrak{G}(n, \mathfrak{k})$ is equivalent to a tensor representation.*[2]

PROOF: Our procedure is to first determine the number of irreducible modular representations of $\mathfrak{G}$ and then to prove the existence of that many distinct irreducible tensor representations.

Brauer and Nesbitt [1. P. 562] have shown that the number of irreducible modular representations of any group is equal to the number of classes of conjugate elements with order prime to p. Now a $\mathfrak{k}$-matrix A is of order prime to p if and only if it is similar to a diagonal matrix (perhaps in some extension of $\mathfrak{k}$); i.e. if and only if the minimum equation of A has no repeated roots. Now for any $\mathfrak{k}$-polynomial of degree n with leading coefficient unity and constant term different from zero there is just one class of similar matrices A with this polynomial as characteristic equation and minimum equation without repeated roots. The number of such polynomials is $q^{n-1}(q - 1)$ where q is the number of elements in $\mathfrak{k}$. The above argument thus shows that *there are exactly $q^{n-1}(q - 1)$ irreducible modular representations of $\mathfrak{G}$.*

[2] See footnote 1.

In the second part of the proof we employ the notation and constructions of the preceding section. We proved there that every Schur function $\{\lambda\}$ is the character of some modular representation of $\mathfrak{G}$. If the eigenvalues of A are $\epsilon^{\alpha_1}, \epsilon^{\alpha_2}, \cdots, \epsilon^{\alpha_n}$ then $\{\lambda(A)\}$ is a symmetric function of $w^{\alpha_1}, w^{\alpha_2}, \cdots, w^{\alpha_n}$. Now every elementary symmetric function of n variables can be expressed as a linear combination of the $\{\lambda\}$ with integer coefficients, and conversely each is an integral linear combination of power products of the elementary symmetric functions [3. Chapter VI]. If we consider the values of all the functions $\{\lambda\}$ only for arguments of the form $w^{\alpha_1}, w^{\alpha_2}, \cdots, w^{\alpha_n}$ that come from an A in $\mathfrak{G}$ there will be certain linear relations, with rational integer coefficients, connecting them. Let

$$\sum_{(\lambda)} a^0(\lambda)\{\lambda(A)\} = 0 \tag{56}$$

be any such linear relation holding for all A in $\mathfrak{G}$. We lose no generality in supposing that the $a^0(\lambda)$ are rational integers not all divisible by p. If we express the $\{\lambda\}$ in terms of the elementary symmetric functions $E_1, \cdots, E_n$ of $w^{\alpha_1}, \cdots, w^{\alpha_n}$ we get from (56) a relation of the form

$$\sum_{(j)} c^0(j)E_1^{j_1}\cdot \cdots \cdot E_n^{j_n} = 0 \tag{57}$$

where again the $c^0(j)$ are rational integers not all divisible by p. All the terms in (57) are integers in $\mathfrak{k}^0$ and so by taking residue classes modulo that prime ideal which maps the integers of $\mathfrak{k}^0$ into $\mathfrak{k}$ we get

$$\sum_{(j)} c(j)e_1^{j_1}\cdot \cdots \cdot e_n^{j_n} = 0 \tag{58}$$

where the $c(j)$ are elements of the prime field $GF(p)$ in $\mathfrak{k}$, not all zero; and the e_i's are the elementary symmetric functions of $\epsilon^{\alpha_1}, \cdots, \epsilon^{\alpha_n}$.

Hence the number of (integrally) linearly independent characters $\{\lambda(A)\}$ is at least as great as the number of monomials in $e_1, \cdots, e_n$ that are linearly independent over the $GF(p)$, where $e_1, \cdots, e_n$ range over $\mathfrak{k}$ subject only to the restriction $e_n \neq 0$. Now there are $q^{n-1}(q-1)$ monomials of degree less than q in $e_1, \cdots, e_{n-1}$, and of degree less than $q - 1$ in e_n, and these monomials are even independent over $\mathfrak{k}$. (For $i < n$ we must allow all exponents less than q for e_i, but any exponents exceeding $q - 1$ can be reduced using the identity $e_i^q = e_i$. For $i = n$ the condition $e_n \neq 0$ enables us to reduce all exponents below $q - 1$ by application of the identity $e_n^{q-1} = 1$. The independence of the resulting monomials is covered by the discussions in Part I §3, see also §14 below.)

Now the number of distinct irreducible tensor representations is certainly at least as great as the number of independent characters in any set of tensor

representations, in particular it must be at least as great as the number $q^{n-1}(q-1)$ of independent $\{\lambda\}$. But we saw above that the total number of modular irreducible representations was $q^{n-1}(q-1)$ and so the theorem is proved.

§12. General program for calculation of modular tensor characters

Consider an indecomposable (direct) constituent $\mathfrak{U}$ of the tensor representation $A \to \Pi_m(A)$ of $\mathfrak{G}$($\mathfrak{k}$ now a field of characteristic p with at least m elements in it, and $\mathfrak{k}^0$ a $\mathfrak{p}$-adic closed field which induces $\mathfrak{k}$). The space for $\mathfrak{U}$ has the form $e\mathfrak{V}_m$ where e is a primitive idempotent of the $\mathfrak{k}$-group ring $\mathfrak{R}$ of $\mathfrak{S}_m$. Let e^0 be an idempotent counterimage of e in $\mathfrak{R}^0$ [5. P. 368]. In general e^0 is not primitive, so suppose $e^0 = e_1^0 + e_2^0 + \cdots$ where e_i^0 is associated with the partition (λ^i) of m. (So far as we now know several of the e_i^0 may be associated with the same partition.) Now $e^0\mathfrak{V}_m^0$ is space for a representation $\mathfrak{U}^0$ which induces $\mathfrak{U}$ in the same sense that $\mathfrak{H}^0(\lambda)$ induces $\mathfrak{H}(\lambda)$. The irreducible constituents of $\mathfrak{U}$ will therefore be the same as those of the representation

$$\mathfrak{U}' = \left\| \begin{matrix} \mathfrak{H}(\lambda^1) & & & \\ & \mathfrak{H}(\lambda^2) & & \\ & & \cdot & \\ & & & \cdot \\ & & & & \cdot \end{matrix} \right\|.$$ Hence the character of $\mathfrak{U}$

will be $\{\lambda^1\} + \{\lambda^2\} + \cdots$. Thus the characters of the indecomposable constituents of the tensor representation π_m are known as soon as the decomposition numbers of $\mathfrak{S}_m$ are known [1. P. 558; 7. Pp. 657–8]. At present those are known only for $m < 2p$.

To calculate the characters of the irreducible constituents of Π_m requires knowledge of the irreducible constituents of each indecomposable direct constituent $\mathfrak{U}$, and in general probably further information as well.

§13. Modular tensor characters for $m < 2p$

For $m < p$ $\mathfrak{R}(= \mathfrak{R}(m, \mathfrak{k}))$ is a semi-simple algebra and so each counterimage idempotent e^0 of a primitive idempotent e is again primitive, and each indecomposable constituent $\mathfrak{U}$ of Π_m is irreducible. Hence the $\mathfrak{H}(\lambda)$ are the irreducible (and indecomposable) constituents of π_m; and thus all characters are known.

For $p \leqq m < 2p$ the nature of the $\mathfrak{U}$'s is given in Part I (Formula (50) and Theorem VII). If (λ) is a partition without a hook of length p [7. Pp. 665–6; 5. Part I], then (in the notation of Theorem VII) [8. P. 683] $\mathfrak{G}(\lambda) = \mathfrak{H}(\lambda)$ is an irreducible, indecomposable direct constituent of Π_m, and so its character is the Schur function $\{\lambda\}$.

The remaining partitions of m are grouped together in sets of p according to the partitions (μ) of $m - p$ obtained by removing the hook of length p, those

belonging to a given (μ) being labelled $\lambda^1(\mu), \cdots, \lambda^p(\mu)$ according to the height [7. P. 665] of the hook. Then from formula (50) we get

$$(59) \qquad \chi_{\mathfrak{U}^i(\mu)} = \chi_{\mathfrak{G}_{i-1}(\mu)} + 2\chi_{\mathfrak{G}_i(\mu)} + \chi_{\mathfrak{G}_{i+1}(\mu)}, \qquad i = 1, \cdots, p-1,$$

and from the decomposition numbers for $\mathfrak{S}_m$:

$$(60) \qquad \chi_{\mathfrak{U}^i(\mu)} = \{\lambda^i(\mu)\} + \{\lambda^{i+1}(\mu)\}, \qquad i = 1, \cdots, p-1.$$

(We recall the conventions of Part I for discarding non-existent parts which appear formally [8. P. 683].)

By Lemma IX and the argument preceding it we know that the irreducible constituents of the $\mathfrak{H}(\lambda)$ are irreducible constituents of Π_m. Equations (59) and (60) show that $\mathfrak{H}(\lambda^i(\mu))$ can have as irreducible constituents only $\mathfrak{G}$'s belonging to (μ). Thus we can write

$$(61) \qquad \{\lambda^i(\mu)\} = \sum_{j=0}^{j=p-1} \alpha_{ij}\chi_{\mathfrak{G}_j(\mu)}, \qquad i = 1, \cdots, p,$$

for non-negative integers $\alpha_{ij}\ (= \alpha_{ij}(\mu))$. Substituting (61) in (60) and equating coefficients of the $\mathfrak{G}$'s in the two expressions (59) and (60) we get

$$\alpha_{ij} + \alpha_{i+1\,j} = \begin{matrix} 1 \text{ if } j = i-1 \\ 2 \text{ if } j = i \\ 1 \text{ if } j = i+1 \\ 0 \text{ otherwise} \end{matrix} \qquad i = 1, \cdots, p.$$

Hence $\alpha_{ij} = 0$ unless $j = i$ or $j = i - 1$, and then it follows that $\alpha_{ii-1} = \alpha_{ii} = 1$, in other words we have

$$(62) \qquad \{\lambda^i(\mu)\} = \chi_{\mathfrak{G}_{i-1}(\mu)} + \chi_{\mathfrak{G}_i(\mu)}, \qquad i = 1, \cdots, p,$$

and the other way around

$$(63) \qquad \chi_{\mathfrak{G}_i(\mu)} = \{\lambda^{i+1}(\mu)\} - \{\lambda^{i+2}(\mu)\} + \cdots + (-1)^i\{\lambda^p(\mu)\}, \qquad i = 0, \cdots, p-1.$$

These last two formulae can be used to compute the characters of Kronecker products of tensor representations, provided the total rank never reaches $2p$. The following theorem summarizes the above results.

THEOREM IX. *Suppose $m < 2p$ and that $\mathfrak{k}$ has at least m elements. If (λ) is a partition of m without any hook of length p, then $\mathfrak{H}(\lambda)$ is an irreducible direct constituent of Π_m. If (μ) is a partition of $l = m - p$ then $\mathfrak{H}(\lambda^i(\mu))$ has the irreducible constituents $\mathfrak{G}_{i-1}(\mu)$ and $\mathfrak{G}_i(\mu)$, each with multiplicity one.*

§14. Existence of a faithful tensor representation of the group ring of the full linear group

Let $\mathfrak{k}$ be a field of characteristic p and of finite order q. We are interested in the $\mathfrak{k}$-group ring $\Gamma = \Gamma(n, \mathfrak{k})$ of the full linear group $\mathfrak{G}(n, \mathfrak{k})$. The $\mathfrak{k}$-enveloping

algebra of any representation of $\mathfrak{G}$ is of course a representation of Γ. In particular, the $\mathfrak{k}$-enveloping algebra $\mathfrak{A}_m$ of Π_m affords a representation of Γ.

THEOREM X. *Let* $c = n^2(q - 1)$. *Then* $\Gamma(n, \mathfrak{k})$ *is isomorphic to the direct sum* $\mathfrak{A}_0 + \cdots + \mathfrak{A}_c$ *of its representations* A_m.

PROOF: In Part I (§§3, 4) we saw that the order of $\mathfrak{A}_m$ is the number of linearly independent forms in $N = n^2$ variables y_{ik} (whose domain is $\mathfrak{k}$) no non-trivial linear combination of which is annihilated upon multiplication by $D = \det \| y_{ik} \|$. Similarly the order, t, of the direct sum $\mathfrak{A}_0 + \cdots + \mathfrak{A}_c$ is the number of linearly independent polynomials of degree c or less in the y_{ik} no non-trivial linear combination of which is annihilated by D. To prove our theorem it is sufficient to show that t equals the order of Γ; for if the homomorphic image of an algebra has order equal to that of the algebra then the homomorphism is an isomorphism.

The set of all $\mathfrak{k}$-polynomials in N variables y_i whose domain is $\mathfrak{k}$ is a commutative $\mathfrak{k}$-algebra $\mathfrak{J}$. By means of the identity $y^q = y$ which holds for all elements of $\mathfrak{k}$ we can replace any polynomial P in the y_i by a polynomial P^* in which no exponent exceeds $q - 1$. If P already has all exponents less than q we call it a *reduced polynomial* and in general call P^* the *reduced* form of P. Thus we see that the order of $\mathfrak{J}$ cannot exceed q^N, which is the number of reduced monomials in N variables.

To show that the order of $\mathfrak{J}$ is actually equal to q^N, we exhibit the q^N polynomials

$$e(a) = \prod_{i=1}^{N} (1 - (y_i - a_i)^{q-1}),$$

one for each N-vector $(a) = (a_1, \cdots, a_N)$ over $\mathfrak{k}$. Unless $y_i = a_i$, $i = 1, \cdots, N$, we have $e(a) = 0$; and if $y_i = a_i$, $i = 1, \cdots, N$, we have $e(a) = 1$. The $e(a)$ are clearly linearly independent, and so form a basis for $\mathfrak{J}$. It is easy to see that the $e(a)$ are orthogonal idempotents, and this shows that $\mathfrak{J}$ is a semi-simple algebra, being the direct sum of q^N copies of $\mathfrak{k}$.

If P is any element of $\mathfrak{J}$ then the set of all Q in $\mathfrak{J}$ which annihilate P form a two-sided ideal denoted by $\mathfrak{J}(D)$. The *element* $P' = 1 - P^{q-1}$ *is a generating idempotent for this ideal.* To prove this we note first that P' is an idempotent annihilator of P and second that if $QP = 0$ then $QP' = Q - QP^{q-1} = Q$.

Returning now to our case of $N = n^2$ variables y_{ik}, we are interested in the annihilating ideal $\mathfrak{J}(D)$. It will have $\mathfrak{k}$-order s where s is the number of the idempotents $e(a_{ik})$ which belong to $\mathfrak{J}(D)$. Now $De(a_{ik})$ has value $\det \| a_{ik} \|$ if $y_{ik} = a_{ik}$ (for all i and k) and value zero for all other choices of y_{ik}. Hence, s equals the number of singular n-rowed $\mathfrak{k}$-matrices, and $q^{n^2} - s$ will thus be the number $(q^n - 1)(q^n - q) \cdot \cdots \cdot (q^n - q^{n-1})$ of non-singular n-rowed $\mathfrak{k}$-matrices. That is, $q^{n^2} - s$ is the order of $\mathfrak{G}(n, \mathfrak{k})$ and therefore of $\Gamma(n, \mathfrak{k})$.

Now to complete the proof of our theorem we observe that $c = n^2(q - 1)$ is the maximum degree of any reduced monomial in n^2 variables and so the number, t, of linearly independent polynomials of degree c or less in n^2 variables,

no linear combination of which lies in $\mathfrak{J}(D)$ is just $q^{n^2} - s$, i.e. equals the order of Γ.

Theorem X of course implies Theorem VIII, but we keep the independent proof of that theorem because of other information which it gives about the irreducible representations.

§15. More about $\mathfrak{J}(D)$, especially for $n = 2$

The proof of Theorem X points the way to the study of each $\mathfrak{A}_m$. We shall call a polynomial P in the y_{ik} a *semiform* if $P = Q^*$ where Q is a form. Then the order of $\mathfrak{A}_m$ is the number $N(n^2, q, m)$ of linearly independent semiforms of degrees $m, m - q + 1, m - 2(q - 1), \cdots$ minus the number $D(n^2, q, m)$ of linearly independent forms of these same degrees which lie in the ideal $\mathfrak{J}(D)$. (In the sequence of degrees $m - i(q - 1)$ we stop with the term for which $1 \leqq m - i(q - 1) < q$.) In the following section we compute $N(n^2, q, m)$ for all $m < 2q$; and in the present section we study $D(n^2, q, m)$ for $m < 2q$ with especial attention (including additional material) for the case $n = 2$.

One important question, as yet unsolved for the general n is the lowest degree $m = M$ for which $D(n^2, q, M) \neq 0$. The relation $y_{11}(D^{q-1} - 1)D = 0$ shows that $M \leqq n(q - 1) + 1$. It can be shown (by a rather direct but clumsy and uninteresting argument which we omit) that for $n > 2$, $M > 2q - 1$. For $n = 2$ we shall now show that $M = 2q - 1$, and shall moreover give a complete determination of the structure of $\mathfrak{J}(D)$ for this case.

We recall that the order of the ideal $\mathfrak{J}(D)$ is the number s of singular $\mathfrak{k}$-matrices; which is, for $n = 2$, $q^3 + q^2 - q$. It is clear that we can select a $\mathfrak{k}$-basis for $\mathfrak{J}(D)$ from products of its idempotent generator $D' = 1 - D^{q-1}$ by monomials. We now show that there are only s distinct products of D' by monomials; it will then follow that these are a $\mathfrak{k}'$-basis for $\mathfrak{J}(D)$.

We associate with the monomial $M(\alpha) = \Pi y_{ik}^{\alpha(ik)}$ the matrix $(\alpha) = \| \alpha(ik) \|$ of exponents, the $\alpha(ik)$ being rational integers less than q. By repeated applications of the identity

$$y_{11}y_{22}D' = y_{12}y_{21}D' \tag{64}$$

(obtained from $DD' = 0$) we see that each product $M(\alpha)D'$ is equal to some product $M(\beta)D'$ where $\beta(12)\ \beta(21) = 0$; and so henceforth we consider only monomials $M(\alpha)$ with $\alpha(12)\ \alpha(21) = 0$. Next, if $\alpha(21) = 0$ and $\alpha(12) + \min(\alpha(11), \alpha(22)) \geqq q$ we can use the identity (64) to replace $M(\alpha)$ by a monomial $M(\beta)$ of lower degree and still have $M(\alpha)D' = M(\beta)D'$. This new (β) might have $\beta(12)\ \beta(21) \neq 0$; if so we can again apply (64) to get $M(\gamma)$ of the same degree as $M(\beta)$, with $\gamma(12)\ \gamma(21) = 0$, and with $M(\gamma)D' = M(\alpha)D'$. A similar argument applies if $\alpha(12) = 0$. So we are justified in henceforth restricting ourselves to monomial multiplers $M(\alpha)$ with

$$\alpha(12)\ \alpha(21) = 0; \quad \alpha(12) + \alpha(21) + \min(\alpha(11), \alpha(22)) < q. \tag{65}$$

There are still further eliminations possible, but for these we must study the form of D' somewhat closer. Since $D^{q-1} = D^q/D$ we have

$$D^{q-1} = y_{11}^{q-1}y_{22}^{q-1} + y_{11}^{q-1}y_{22}^{q-2}y_{12}y_{21} + \cdots + y_{12}^{q-1}y_{21}^{q-1}.$$

From this we see that if $\alpha(11)\ \alpha(22) \neq 0$, the first two terms of $M(\alpha)D'$ cancel leaving the $q - 1$ terms

$$M(\alpha)D' = -\sum_{\nu=1}^{q-1} y_{11}^{q-\nu-1+\alpha(11)}\ y_{22}^{q-\nu-1+\alpha(22)}\ y_{12}^{\nu+\alpha(12)}\ y_{21}^{\nu+\alpha(21)},$$

where exponents exceeding $q - 1$ are to be reduced by $q - 1$. The terms in this sum form a cycle, each term being obtained from the preceding by subtracting 1 from the exponents of y_{11}, y_{22} and adding 1 to the exponents of y_{12}, y_{21}, except that when subtraction would leave a zero exponent we put in $q - 1$ and when addition would give q we put in 1. (In particular this rule applies from the last term to the first, hence the word "cycle.") Thus the whole cycle is determined by any one of its terms, and two such cycles will either be identical or have no term in common. We can therefore characterize the whole sum $M(\alpha)D'$ by its term involving y_{11}^{q-1}, i.e. the one with $\nu = \alpha(11)$:

$$y_{11}^{q-1}\ y_{22}^{q-1+\alpha(22)-\alpha(11)}\ y_{12}^{\alpha(11)+\alpha(12)}\ y_{21}^{\alpha(11)+\alpha(21)}, \tag{66}$$

where some of the exponents may have to be reduced by $q - 1$.

It is clear that no two distinct monomials both with $\alpha(12) = 0$ or both with $\alpha(21) = 0$ can give rise to cyclic products having the same term (66). However, if $\alpha(12) = 0$, $\beta(21) = 0$, and

$$\begin{aligned} &\alpha(22) - \alpha(11) \equiv \beta(22) - \beta(11), \alpha(11) \equiv \beta(11) + \beta(22), \\ &\alpha(11) + \alpha(21) \equiv \beta(11) \bmod (q - 1) \end{aligned} \tag{67}$$

then $M(\alpha)D' = M(\beta)D'$. The congruences (67) set up a one to one correspondence between all monomials $M(\alpha)$ with $\alpha(11)\ \alpha(22)\ \alpha(21) \neq 0$, $\alpha(12) = 0$, and all monomials $M(\beta)$ with $\beta(11)\ \beta(22)\ \beta(12) \neq 0$, $\beta(21) = 0$. For our calculation of the number of distinct products $M(\alpha)D'$ we need to know for how many corresponding pairs (α) and (β) defined by (67) do both (α) and (β) satisfy the second condition of (65)? For the sake of definiteness in choosing a basic set of monomials we will agree to keep the one in such a pair which has $\beta(21) = 0$, and discard the other. Solving for (β) in (67) we get

$$\begin{aligned} &\beta(11) \equiv \alpha(11) + \alpha(22) \bmod (q - 1), \quad \beta(22) \equiv \alpha(22) + \alpha(21) \bmod (q - 1), \\ &\beta(12) = q - 1 - \alpha(21). \end{aligned} \tag{68}$$

First case: If both $\alpha(11) + \alpha(21) < q$ and $\alpha(22) + \alpha(21) < q$ then the congruences in (68) can be replaced by equalities. Hence $\beta(11) + \beta(12) = q - 1 + \alpha(11)$ and $\beta(22) + \beta(12) = q - 1 + \alpha(22)$ are both greater than $q - 1$ so that (β) does not satisfy (65); by our agreement, all such (α) must be kept.

Second case: If $\alpha(11) + \alpha(21) \geqq q$ then $\beta(11) = \alpha(11) + \alpha(21) - q + 1$ and so (β) satisfies (65). Similarly if $\alpha(22) + \alpha(21) \geqq q$ then (β) satisfies (65). In both instances (α) must be rejected.

The cases discussed exhaust all possibilities; so we are now ready to collect the above results and count the remaining monomials. In the following matrices no $\alpha(ik)$ which appears is permitted to take the value zero:

$$
\begin{aligned}
&\text{(i)} \quad \begin{Vmatrix} 0 & 0 \\ 0 & 0 \end{Vmatrix}; \qquad \text{(ii)} \quad \begin{Vmatrix} \alpha(11) & 0 \\ 0 & 0 \end{Vmatrix}, \begin{Vmatrix} 0 & \alpha(12) \\ 0 & 0 \end{Vmatrix}, \begin{Vmatrix} 0 & 0 \\ \alpha(21) & 0 \end{Vmatrix}, \begin{Vmatrix} 0 & 0 \\ 0 & \alpha(22) \end{Vmatrix}; \\
&\text{(iii)} \quad \begin{Vmatrix} \alpha(11) & \alpha(12) \\ 0 & 0 \end{Vmatrix}, \begin{Vmatrix} \alpha(11) & 0 \\ \alpha(21) & 0 \end{Vmatrix}, \begin{Vmatrix} 0 & \alpha(12) \\ 0 & \alpha(22) \end{Vmatrix}, \begin{Vmatrix} 0 & 0 \\ \alpha(21) & \alpha(22) \end{Vmatrix}; \\
(69) \quad &\text{(iv)} \quad \begin{Vmatrix} \alpha(11) & 0 \\ 0 & \alpha(22) \end{Vmatrix}; \\
&\text{(v)} \quad \begin{Vmatrix} \alpha(11) & 0 \\ \alpha(21) & \alpha(22) \end{Vmatrix}, \quad \alpha(21) + \min(\alpha(11), \alpha(22)) < q; \\
&\text{(vi)} \quad \begin{Vmatrix} \alpha(11) & \alpha(12) \\ 0 & \alpha(22) \end{Vmatrix}, \quad \alpha(12) + \max(\alpha(11), \alpha(22)) < q.
\end{aligned}
$$

There are $1, 4(q-1), 4(q-1)^2, (q-1)^2$ matrices of types (i), (ii), (iii), (iv) respectively. To calculate the number, N_1, of type (v) we observe that if $\alpha(21) = j$, then there are $(q-1)^2 - j^2$ pairs of values for $\alpha(11)$ and $\alpha(22)$ for which $\alpha(21) + \min(\alpha(11), \alpha(22)) < q$ and so

$$N_1 = \sum_{j=1}^{q-1} (q-1)^2 - j^2.$$

Similarly the number, N_2, of type (vi) turns out to be

$$N_2 = \sum_{j=1}^{q-1} (q - 1 - j)^2 = \sum_{j=1}^{q-2} j^2.$$

Hence $N_1 + N_2 = (q-1)^3 - (q-1)^2$. Adding all the above numbers together we get $q^3 + q^2 - q = s$, which was our goal.

Therefore, the above set of s matrices $\| \alpha(ik) \|$ give s products $M(\alpha)D'$ which are a $\mathfrak{k}$-basis for $\mathfrak{J}(D)$. If $\| \alpha(ik) \|$ is one of the above matrices we shall refer to $M(\alpha)$ as a *selected monomial.* Note that no selected monomial has degree exceeding $2q - 2$.

This completes the general study of the ideal $\mathfrak{J}(D)$, but for application to the study of $\mathfrak{A}_{2q-1}$ a little further analysis is needed.

LEMMA X. *There are exactly* $4q$ *linearly independent semiforms of degree* $2q - 1$ *in the ideal* $\mathfrak{J}(D)$.

PROOF: Any semiform of degree $2q - 1$ is a sum of monomials of degree 1, q, and $2q - 1$. There are 4 selected monomials of degree 1, and $4(q - 1)$ selected monomials of degree q and of type (iii); for each of these monomials the product with D' has degree $2q - 1$. The remaining selected monomials of degree q have

$\alpha(11)\ \alpha(22) \neq 0$ and so their products with D' have the cyclic character described above: no two having a term in common, and all of degree $3q - 2$. Hence no $\mathfrak{k}$-combination of such products can have degree $2q - 1$. Since there are no selected monomials of degree $2q - 1$ our proof is now complete.

§16. The calculation of $N(n^2, q, m)$ for $q \leqq m < 2q$

The number $N(N, q, m)$ of linearly independent semiforms of degrees m, $m - q + 1, \cdots$ in N variables y_i is clearly equal to the total number of reduced monomials of these degrees. For a monomial M of degree $m < 2q$ the reduced monomial M can differ from M in at most one variable. In the following argument we first assume $q \leqq m < 2q - 1$ and then supply a modified argument for the remaining case $m = 2q - 1$.

If $M^* \neq M$, then M^* has degree $l = m - q + 1$, so $N(N, q, m) =$ number of monomials of degree l plus the number of reduced monomials of degree m. Let M^* be a monomial of degree l involving exactly β_0 variables; β_1 to the first degree, β_2 to the second degree, $\cdots$, β_i to the i^{th} degree, $\cdots$. Then $\beta_0 = \beta_1 + \beta_2 \cdots + \beta_l$, and $l = \beta_1 + 2\beta_2 + \cdots + l\beta_l$. M^* is the reduced monomial of exactly β_0 monomials M of degree m, namely, those obtained by adding $q - 1$ to the exponent of each variable in turn. There are

$$\binom{N}{\beta_0}\binom{\beta_0}{\beta_1}\binom{\beta_0 - \beta_1}{\beta_2}\cdots\binom{\beta_0 - \beta_1 - \cdots \beta_{l-1}}{\beta_l} = \binom{N}{\beta_0}\frac{\beta_0!}{\beta_1!\beta_2!\cdots\beta_l!}$$

different monomials of degree l having the same exponents (aside from arrangement) as M^*. Every non-reduced M can be obtained in this way for some partition (β) of l. Hence, if we set

$$R(N, q, l) = \sum_{(\beta)} (\beta_0 - 1)\binom{N}{\beta_0}\frac{\beta_0!}{\beta_1!\beta_2!\cdots\beta_l!} \tag{70}$$

the sum being over all partitions (β) of l, we have that $N(N, q, m) + R(N, q, l) = \binom{N + m - 1}{m} =$ total number of monomials of degree m in N variables. We now calculate $R(N, q, l)$.

LEMMA XI. $R(N, q, l) = (l - 1)\binom{N + l - 2}{l}$.

This lemma is the case $j = 0$ of the following formula:

$$(l - 1)\binom{N + l - j - 2}{l - j} = \sum_{(\beta)} (\beta_0 - 1)\binom{N - j}{\beta_0 - j}\frac{\beta_0!}{\beta_1!\beta_2!\cdots\beta_l!}. \tag{71}$$

This formula is readily verified for small N; for any N and $j = l - 1$ it becomes the identity

$$(l - 1)\binom{N - 1}{1} = (l - 1)\binom{N - l + 1}{1} + (l - 2)\binom{N - l + 1}{0}\frac{(l - 1)!}{(l - 2)!1!},$$

since there are no terms on the right with $\beta_0 < j$ and there are only two partitions of l with $\beta_0 \geqq j = l - 1$. Now we proceed by induction; i.e. we consider a particular N and j and take as induction hypothesis that (71) is valid for all smaller N, and for the given N with all larger j. Apply the general binomial relation $\binom{r}{s} = \binom{r-1}{s} + \binom{r-1}{s-1}$ to the left side of (71) and we get two terms which are the left sides of (71) for $N - 1$ and j, and for N and $j + 1$. The same happens on the right hand side, and this completes the induction argument.

The above computations were based on the fact that for $M \neq M^*$ the difference in degrees was $q - 1$. For $m = 2q - 1$ the difference in degrees can be either $q - 1$ or $2q - 2$, but the latter arises only for the N monomials y_i^{2q-1} and since they give rise to N distinct reduced monomials, we can apply Lemma XI for $l = q$ to obtain the number of independent "relations" between monomials of degree $2q - 1$.

§17. Preparations for the reduction of Π_m for $p \leqq m < 2p - 1$, and $q = p$

The number $R(n^2, p, m - p + 1)$ calculated in the previous section can be thought of as the loss in the order of $\mathfrak{A}_m$ due to replacing a $\mathfrak{k}$ with more than p elements by the $\mathfrak{k}$ with p elements. We now show how to account for this loss of order in terms of the representations.

Our procedure is first to present Π_m in the reduced form given by Theorem VII (which is, of course, valid for all $\mathfrak{k}$), and then to consider what further reductions are possible when $\mathfrak{k}$ has just p elements. We will need to study the case of general $\mathfrak{k}$ in more detail than is given by Theorem VII; to minimize the shuttling back and forth from one $\mathfrak{k}$ to another we will first do all the work for the larger $\mathfrak{k}$ and then shift to the smaller one.

Since $a \to a^p$ is a homomorphism (automorphism if $\mathfrak{k}$ is finite) of $\mathfrak{k}$ into itself the mapping $\| a_{ik} \| \to \| a_{ik}^p \|$ is a representation of the full linear group $\mathfrak{G}$.

For the rest of this section we suppose that $\mathfrak{k}$ has $q > p$ elements. Then the character of the above representation of $\mathfrak{G}$ is readily computed as s_p = sum of the p^{th} powers of the complex number counterimages of the modular eigenvalues of $\| a_{ik} \|$. Now we have [3. P. 86] $s_p = \{p\} - \{p - 1, 1\} + \cdots + \{1^p\}$. For $m = p$, we call $(0): 0 = 0$ a partition of zero, and so in the notation of Theorem VII we have $(p - i, 1^i) = \lambda^{i+1}(0)$. Then from formula (63) we get

$$\chi_{\mathfrak{G}_0(0)} = \{\lambda^1(0)\} - \{\lambda^2(0)\} + \cdots + \{\lambda^p(0)\} = s_p. \tag{72}$$

Applying the Brauer-Nesbitt criterion that two irreducible representations are equivalent if their characters are equal we see that (with proper choice of basis elements) if $A = \| a_{ik} \|$ then $\mathfrak{G}_0(0)(A) = \| a_{ik}^p \|$.

Lemma XII. *If $\mathfrak{k}$ has more than p elements and $l \leqq p$, the enveloping algebra $\mathfrak{H}$ of $\mathfrak{G}_0(0) \times \Pi_{l-1}$ has order h equal to the product of the orders of the factors, i.e.*

$$h = n^2 \binom{n^2 + l - 2}{l - 1}.$$

Furthermore, the Kronecker product of $\mathfrak{S}_0(0)$ by any irreducible constituent $\mathfrak{S}(\mu)$ of Π_{l-1} is the irreducible constituent $\mathfrak{S}_0(\mu)$ of $\mathfrak{H}$ and $\mathfrak{H}$ is the direct sum of such constituents.

PROOF: The elements of Π_{l-1} are monomials of degree $l - 1$ in the elements a_{ik} of A. Hence, if the order of the Kronecker product were less than h there would have to be a non-zero form $P(b_{ik}, a_{ik})$, each term linear in the variables b_{ik} and of degree $l - 1$ in the variables a_{ik}, such that $P(a_{ik}^p, a_{ik}) = 0$ for all values of the a_{ik}. Since $m = p + l - 1 < p^2 \leqq q$ this means that $P(A_{ik}^p, A_{ik})$ must vanish identically as a form in n^2 indeterminants A_{ik}. But this cannot happen, since $A_{ik}^p M = A_{i'k'}^p M'$ with M and M' monomials of degree $l - 1 < p$ requires $i = i'$, $k = k'$, and therefore $M = M'$. In other words $P(A_{ik}^p, A_{ik})$ has exactly the same coefficients as the initial $P(b_{ik}, a_{ik})$. This establishes the value of h.

The same argument can be used to establish the irreducibility of $\mathfrak{S}(\mu) \times \mathfrak{S}_0(0)$ provided we take for M and M' forms of degree $l - 1$. The equality $\mathfrak{S}(\mu) \times \mathfrak{S}_0(0) = \mathfrak{S}_0(\mu)$ follows from the character formula (63), and the relation

$$\{\mu\}\, s_p = \{\lambda^1(\mu)\} - \{\lambda^2(\mu)\} + \cdots + (-1)^{p-1}\{\lambda^p(\mu)\}.$$

This last relation is established by a short computation based on a theorem of Littlewood [3. P. 70, Th. II].

§18. The reduced form of Π_m for $p \leqq m < 2p - 1$, and $q = p$

In this section we take $\mathfrak{f}$ with exactly p elements. Our procedure is first to present Π_m in the reduced form given by Theorem VII (which is, of course, valid for all $\mathfrak{f}$ of characteristic p), and then to consider what further reductions can be made due to the shrinking of the size of $\mathfrak{f}$.

Since $a^p = a$ for all a in $\mathfrak{f}$ we have that $\mathfrak{S}_0(0)(A) = \| a_{ik}^p \| = \| a_{ik} \| = A$; i.e. $\mathfrak{S}_0(0) = \Pi_1$. Hence, $\mathfrak{S}_0(0) \times \Pi_{l-1} = \Pi_l$; and if $l < p$ this is fully reducible of order $h' = \binom{n^2 + l - 1}{l}$. We regard Π_m as being the product $\Pi_p \times \Pi_{l-1}$ and thus see that each irreducible constituent of the $\mathfrak{H}$ of Lemma XII must now decompose into a direct sum of irreducible constituents of Π_l. The irreducible constituents of $\mathfrak{H}$ were irreducible constituents of Π_m and might appear in the product $\Pi_p \times \Pi_{l-1}$ in various places, but wherever they appeared they must now decompose. The actual form of this decomposition can be determined by studying the characters, making use of the relation $s_p = s_1$ which holds for our present $\mathfrak{f}$. (We shall make frequent use of such computations below in the treatment of the case $m = 2$, $m = 2p - 1$, $q = p$.) The difference $h - h'$ is the loss in order of $\mathfrak{A}_m$ due to the relation $\mathfrak{S}_0(0) = \Pi_1$. But $h - h' = R(n^2, p, l)$ which we have already seen to be the total loss in order of $\mathfrak{A}_m$. This shows that the only modifications needed to bring $\mathfrak{A}_m$ in fully reduced form for our case are those described above on the irreducible constituents that appear in $\mathfrak{H}$. (For, because of the independence of the various matrix sets appearing in the reduced form of Π_m in Th. VII any further reductions would entail further loss in order.)

since there are no terms on the right with $\beta_0 < j$ and there are only two partitions of l with $\beta_0 \geqq j = l - 1$. Now we proceed by induction; i.e. we consider a particular N and j and take as induction hypothesis that (71) is valid for all smaller N, and for the given N with all larger j. Apply the general binomial relation $\binom{r}{s} = \binom{r-1}{s} + \binom{r-1}{s-1}$ to the left side of (71) and we get two terms which are the left sides of (71) for $N - 1$ and j, and for N and $j + 1$. The same happens on the right hand side, and this completes the induction argument.

The above computations were based on the fact that for $M \neq M^*$ the difference in degrees was $q - 1$. For $m = 2q - 1$ the difference in degrees can be either $q - 1$ or $2q - 2$, but the latter arises only for the N monomials y_i^{2q-1} and since they give rise to N distinct reduced monomials, we can apply Lemma XI for $l = q$ to obtain the number of independent "relations" between monomials of degree $2q - 1$.

§17. Preparations for the reduction of Π_m for $p \leqq m < 2p - 1$, and $q = p$

The number $R(n^2, p, m - p + 1)$ calculated in the previous section can be thought of as the loss in the order of $\mathfrak{A}_m$ due to replacing a $\mathfrak{k}$ with more than p elements by the $\mathfrak{k}$ with p elements. We now show how to account for this loss of order in terms of the representations.

Our procedure is first to present Π_m in the reduced form given by Theorem VII (which is, of course, valid for all $\mathfrak{k}$), and then to consider what further reductions are possible when $\mathfrak{k}$ has just p elements. We will need to study the case of general $\mathfrak{k}$ in more detail than is given by Theorem VII; to minimize the shuttling back and forth from one $\mathfrak{k}$ to another we will first do all the work for the larger $\mathfrak{k}$ and then shift to the smaller one.

Since $a \rightarrow a^p$ is a homomorphism (automorphism if $\mathfrak{k}$ is finite) of $\mathfrak{k}$ into itself the mapping $\| a_{ik} \| \rightarrow \| a_{ik}^p \|$ is a representation of the full linear group $\mathfrak{G}$.

For the rest of this section we suppose that $\mathfrak{k}$ has $q > p$ elements. Then the character of the above representation of $\mathfrak{G}$ is readily computed as s_p = sum of the p^{th} powers of the complex number counterimages of the modular eigenvalues of $\| a_{ik} \|$. Now we have [3. P. 86] $s_p = \{p\} - \{p - 1, 1\} + \cdots + \{1^p\}$. For $m = p$, we call $(0): 0 = 0$ a partition of zero, and so in the notation of Theorem VII we have $(p - i, 1^i) = \lambda^{i+1}(0)$. Then from formula (63) we get

$$\chi_{\mathfrak{G}_0(0)} = \{\lambda^1(0)\} - \{\lambda^2(0)\} + \cdots + \{\lambda^p(0)\} = s_p. \tag{72}$$

Applying the Brauer-Nesbitt criterion that two irreducible representations are equivalent if their characters are equal we see that (with proper choice of basis elements) if $A = \| a_{ik} \|$ then $\mathfrak{G}_0(0 \wr A) = \| a_{ik}^p \|$.

Lemma XII. *If $\mathfrak{k}$ has more than p elements and $l \leqq p$, the enveloping algebra $\mathfrak{H}$ of $\mathfrak{G}_0(0) \times \Pi_{l-1}$ has order h equal to the product of the orders of the factors, i.e.*

$$h = n^2 \binom{n^2 + l - 2}{l - 1}.$$

Furthermore, the Kronecker product of $\mathfrak{G}_0(0)$ *by any irreducible constituent* $\mathfrak{G}(\mu)$ *of* Π_{l-1} *is the irreducible constituent* $\mathfrak{G}_0(\mu)$ *of* $\mathfrak{H}$ *and* $\mathfrak{H}$ *is the direct sum of such constituents.*

PROOF: The elements of Π_{l-1} are monomials of degree $l - 1$ in the elements a_{ik} of A. Hence, if the order of the Kronecker product were less than h there would have to be a non-zero form $P(b_{ik}, a_{ik})$, each term linear in the variables b_{ik} and of degree $l - 1$ in the variables a_{ik}, such that $P(a_{ik}^p, a_{ik}) = 0$ for all values of the a_{ik}. Since $m = p + l - 1 < p^2 \leqq q$ this means that $P(A_{ik}^p, A_{ik})$ must vanish identically as a form in n^2 indeterminants A_{ik}. But this cannot happen, since $A_{ik}^p M = A_{i'k'}^p M'$ with M and M' monomials of degree $l - 1 < p$ requires $i = i'$, $k = k'$, and therefore $M = M'$. In other words $P(A_{ik}^p, A_{ik})$ has exactly the same coefficients as the initial $P(b_{ik}, a_{ik})$. This establishes the value of h.

The same argument can be used to establish the irreducibility of $\mathfrak{G}(\mu) \times \mathfrak{G}_0(0)$ provided we take for M and M' forms of degree $l - 1$. The equality $\mathfrak{G}(\mu) \times \mathfrak{G}_0(0) = \mathfrak{G}_0(\mu)$ follows from the character formula (63), and the relation

$$\{\mu\}\, s_p = \{\lambda^1(\mu)\} - \{\lambda^2(\mu)\} + \cdots + (-1)^{p-1}\{\lambda^p(\mu)\}.$$

This last relation is established by a short computation based on a theorem of Littlewood [3. P. 70, Th. II].

§18. The reduced form of Π_m for $p \leqq m < 2p - 1$, and $q = p$

In this section we take $\mathfrak{k}$ with exactly p elements. Our procedure is first to present Π_m in the reduced form given by Theorem VII (which is, of course, valid for all $\mathfrak{k}$ of characteristic p), and then to consider what further reductions can be made due to the shrinking of the size of $\mathfrak{k}$.

Since $a^p = a$ for all a in $\mathfrak{k}$ we have that $\mathfrak{G}_0(0)(A) = \| a_{ik}^p \| = \| a_{ik} \| = A$; i.e. $\mathfrak{G}_0(0) = \Pi_1$. Hence, $\mathfrak{G}_0(0) \times \Pi_{l-1} = \Pi_l$; and if $l < p$ this is fully reducible of order $h' = \binom{n^2 + l - 1}{l}$. We regard Π_m as being the product $\Pi_p \times \Pi_{l-1}$ and thus see that each irreducible constituent of the $\mathfrak{H}$ of Lemma XII must now decompose into a direct sum of irreducible constituents of Π_l. The irreducible constituents of $\mathfrak{H}$ were irreducible constituents of Π_m and might appear in the product $\Pi_p \times \Pi_{l-1}$ in various places, but wherever they appeared they must now decompose. The actual form of this decomposition can be determined by studying the characters, making use of the relation $s_p = s_1$ which holds for our present $\mathfrak{k}$. (We shall make frequent use of such computations below in the treatment of the case $m = 2$, $m = 2p - 1$, $q = p$.) The difference $h - h'$ is the loss in order of $\mathfrak{A}_m$ due to the relation $\mathfrak{G}_0(0) = \Pi_1$. But $h - h' = R(n^2, p, l)$ which we have already seen to be the total loss in order of $\mathfrak{A}_m$. This shows that the only modifications needed to bring $\mathfrak{A}_m$ in fully reduced form for our case are those described above on the irreducible constituents that appear in $\mathfrak{H}$. (For, because of the independence of the various matrix sets appearing in the reduced form of Π_m in Th. VII any further reductions would entail further loss in order.)

We have now proved the main theorem:

Theorem XI. *If $\mathfrak{k}$ is the field with p elements and $p \leqq m = p + l - 1 < 2p - 1$ then the reduced form of the Kronecker m^{th} power representation Π_m of the full linear group $\mathfrak{G}(n, \mathfrak{k})$ of non-singular n-rowed $\mathfrak{k}$-matrices is given by Theorem VII except for the following modifications: the diagonal constituent $\mathfrak{G}_0(\mu)$ of Π_m (irreducible for $\mathfrak{k}$ with more than p elements) is now equal to $\mathfrak{G}(\mu) \times \Pi_1$, which is the direct sum $\mathfrak{G}(\mu^1) + \mathfrak{G}(\mu^2) + \cdots$ of irreducible constituents of Π_l; the partitions (μ^1), (μ^2), $\cdots$ being those which appear in the decomposition of the character product $\{\mu\}\{1\} = \{\mu^1\} + \{\mu^2\} + \cdots$.*

§19. The decomposition for $q = p$, $m = 2p - 1$, $n > 2$

Lem. XI and XII both include the case $l = p$, and we know that the order of Π_p is the same for all $\mathfrak{k}$. Hence, it is still true that all the changes due to the smallness of $\mathfrak{k}$ are consequences of the relation $\mathfrak{G}_0(0) = \Pi_1$. However, this time the constituents $\mathfrak{G}_0(\mu) = \mathfrak{G}(\mu) \times \mathfrak{G}_0(0)$ will not always split into a direct sum of irreducible constituents of Π_p. But, since $\mathfrak{G}(\mu)$ is an indecomposable constituent of Π_{p-1}, $\mathfrak{G}(\mu) \times \Pi_1$ will be a direct sum of indecomposable constituents of Π_p, and just which will appear can be determined by the character theory. With these modifications Th. XI will hold for $m = 2p - 1$, $n > 2$. For $n = 2$ these modifications are still valid, but no longer tell the whole story, since we have then (according to Lem. X) to account for a further loss in order of $4p$ due to the presence of determinant annihilators.

§20. The decomposition for $q = p$, $m = 2p - 1$, $n = 2$

For the present section we denote by $\mathfrak{G}$ the full linear group of two rowed matrices for a field with $q > p$ elements and by $\mathfrak{G}^*$ the same for a field with $q = p$ elements. Since $\mathfrak{G}^*$ is a subgroup of $\mathfrak{G}$ any representation $\mathfrak{T}$ of $\mathfrak{G}$ when considered only for elements of $\mathfrak{G}^*$ affords a representation of $\mathfrak{G}^*$ which we denote by $\mathfrak{T}^*$. In particular we apply this notation to the tensor representations and constituents Π_m, $\mathfrak{A}_m$, $\mathfrak{U}^i(\mu)$, $\mathfrak{G}(\lambda)$, $\mathfrak{G}_i(\mu)$, $\mathfrak{G}_i^j(\mu)$ defined above (cf. Th. VII).

Since $n = 2$, we have $\{\lambda\} = 0$ for any partition (λ) of any integer m into more than two summands [10, P. 128, Lem. 4.4B]. Hence in applications of Th. VII for $n = 2$, $q > p$ we must delete all constituents that arise from partitions of m into more than two parts. (Indeed such modifications are always to be made when $n > m$.)

The following string of lemmas, taken with Th. VII, Th. XI, and the results of §19 complete the analysis for $n = 2$, $m = 2p - 1$, $q = p$.

Lemma XIII. *Every partition (λ): $m = \lambda_1 + \lambda_2$ for $m = 2p - 1$ and $m = 2p - 3$ has a hook of length p; indeed, if $\lambda_1 - \lambda_2 \geqq p$ then $(\lambda) = \lambda^1(\lambda_1 - p, \lambda_2)$ and if $\lambda_1 - \lambda_2 < p$ then $(\lambda) = \lambda^2(\lambda_2 - 1, \lambda_1 - p + 1)$.*

We shall use D to stand for det A whether A is in $\mathfrak{G}$ or in $\mathfrak{G}^*$. We recall that for $n = 2$, $D = \{1^2\} = \mathfrak{G}(1^2)$.

Lemma XIV. *If (μ): $p - 3 = \mu_1 + \mu_2$ is any one or two element partition of $p - 3$, and we denote by $(\mu)_D$ the partition $p - 1 = (\mu_1 + 1) + (\mu_2 + 1)$ then*

$\{\lambda^i(\mu)\}\cdot D = \{\lambda^i(\mu)_D\}$, $i = 1, 2$. *Furthermore,* $\lambda^1(p-1)$ *is the only one or two element partition of* $2p-1$ *which does not have the form* $\lambda^i(\mu)_D$.

LEMMA XV. *As* (μ) *runs through all one or two element partitions of* $p - 3$

$$\mathfrak{U}^1(\mu) = \left\| \begin{matrix} \mathfrak{G}_1(\mu) & & \\ \mathfrak{G}_0^1(\mu) & \mathfrak{G}_0(\mu) & \\ \mathfrak{G}_1^1(\mu) & \mathfrak{G}_1^0(\mu) & \mathfrak{G}_1(\mu) \end{matrix} \right\|, \qquad \mathfrak{U}^2(\mu) = \mathfrak{G}_1(\mu)$$

give the indecomposable constituents of Π_{2p-3}. *The* $\mathfrak{U}^i(\mu)^*$ *are the indecomposable constituents of* Π_{2p-3}^*. *The decomposed form of* $\mathfrak{U}^i(\mu)^*$ *differs from that of* $\mathfrak{U}^i(\mu)$ *only in that* $\mathfrak{G}_0(\mu)^* = \mathfrak{G}(\mu)^* \times \Pi_1^*$ *is a direct sum of (at most two) irreducible representations of* $\mathfrak{G}^*$. *In particular, for each* (μ) *the enveloping algebras of* $\mathfrak{U}^1(\mu)$ *and* $\mathfrak{U}^1(\mu)^*$ *have radicals of equal order.*

LEMMA XVI. *For every one or two element partition* (μ) *of* $p - 3$ *we have* (i) $\mathfrak{U}^i(\mu)_D = \mathfrak{U}^i(\mu) \times D$; (ii) $\mathfrak{U}^i(\mu)_D^* = \mathfrak{U}^i(\mu)^* \times D$ *is an indecomposable constituent of* Π_{2p-1}^*, *and* (iii) $\mathfrak{A}_{2p-3}^* \times D$ *has the same order as* $\mathfrak{A}_{2p-3}^*$.

LEMMA XVII. $\mathfrak{U}^1(p-1)^* = \mathfrak{U}^1(0)^* = \left\| \begin{matrix} \mathfrak{G}_1(0)^* & & \\ \mathfrak{G}_0^1(0)^* & \mathfrak{G}(1)^* & \\ \mathfrak{G}_1^1(0)^* & \mathfrak{G}_1^0(0)^* & \mathfrak{G}_1(0)^* \end{matrix} \right\|.$

LEMMA XVIII. *Let* $0(\mathfrak{A}_m^*)$ *denote the order of* $\mathfrak{A}_m^*$. *Then* $0(\mathfrak{A}_{2p-1}^*) - 0(\mathfrak{A}_{2p-3}^*) = (p-1)^2$.

LEMMA XIX. *In* $\mathfrak{U}^1(p-2, 1)^*$ *we have*

(i) $\mathfrak{G}_1(p-2, 1)^* = \mathfrak{G}(1)^*$;

(ii) $\mathfrak{G}_0(p-2, 1)^* = \left\| \begin{matrix} \mathfrak{G}_1(0)^* & \\ & \mathfrak{G}(\nu+1, \nu)^* \end{matrix} \right\|$, *where* $2\nu = p - 1$;

(iii) $\mathfrak{G}_0^1(p-2, 1)^* = \left\| \begin{matrix} \mathfrak{G}_1^0(0)^* \\ * \end{matrix} \right\|$;

(iv) $\mathfrak{G}_1^0(p-2, 1)^* = \| \mathfrak{G}_0^1(0)^* \quad * \|$.

Here, the refinement into submatrices in (iii) *and* (iv) *is coherent with that in* (ii) *and the single "*" denotes a submatrix in whose form we are not now concerned.*

To prove Lemma XIII we need only to observe that $\lambda_1 \geqq \lambda_2$ and $\lambda_1 + \lambda_2 \geqq 2p - 3$ imply that $\lambda_1 \geqq p - 1$.

PROOF OF LEMMA XIV. For $n = 2$ we have $\{\lambda_1, \lambda_2\}\{1^2\} = \{\lambda_1 + 1, \lambda_2 + 1\}$ according to the product rule for Schur functions [3. P. 94, Th. V]. But if $(\lambda_1, \lambda_2) = \lambda^i(\mu)$ then $(\lambda_1 + 1, \lambda_2 + 1) = \lambda^i(\mu)_D$. For the second part of the lemma we note that for any partition (λ'): $2p - 1 = \lambda_1' + \lambda_2'$ with $\lambda_2' \neq 0$ we have $(\lambda') = \lambda^i(\mu)_D$ where (μ) and i are defined by $(\lambda_1' - 1, \lambda_2' - 1) = \lambda^i(\mu)$.

Lemma XV is a restatement, for the present case, of Th. X, taking account of the partition structure for $m = 2p - 3$ given in Lemma XIII.

PROOF OF LEMMA XVI. Since D is a one rowed representation, multiplication of any representation by D cannot change its structure provided no annihilators of D are encountered. But no annihilator of D has degree less than $2p - 2$,

and so $\mathfrak{A}_{2p-3}^* \times D$ has the same order as $\mathfrak{A}_{2p-3}^*$. Furthermore, $\mathfrak{U}^i(\mu) \times D$ and $\mathfrak{U}^i(\mu)^* \times D$ will be indecomposable and will have enveloping algebras isomorphic to those of $\mathfrak{U}^i(\mu)$ and $\mathfrak{U}^i(\mu)^*$, respectively.

Now since $\Pi_{2p-1} = \Pi_{2p-3} \times \Pi_2$ and D is an indecomposable constituent of Π_2, $\mathfrak{U}^i(\mu) \times D$ must be an indecomposable constituent of Π_{2p-1}; then from Lemma XIV and the character formulae of §13 it follows that $\mathfrak{U}^i(\mu) \times D = \mathfrak{U}^i(\mu)_D$. Now put stars on both sides of this equation and we get the remaining assertion of the present lemma.

Proof of Lemma XVII. Since $\lambda^2(p - 1)$ is a three element partition, $\mathfrak{U}^1(p - 1)$ is irreducible and equal to $\mathfrak{G}_0(p - 1)$, which is in turn equal to $\mathfrak{G}(p - 1) \times \mathfrak{G}_0(0)$, (cf. Lem. XII). Hence, $\mathfrak{U}^1(p - 1)^* = \mathfrak{G}(p - 1)^* \times \mathfrak{G}_0(0)^* = \mathfrak{G}(p - 1)^* \times \mathfrak{G}(1)^*$. Now by the argument made in §19, the character relation $\{p - 1\}\{1\} = \{p\} + \{p - 1, 1\}$, and the formulae of §13 we get $\mathfrak{U}^1(p - 1)^* = \mathfrak{U}^1(0)^*$, of which the detailed form is given by Th. VII.

Proof of Lemma XVIII. In §15 we showed that $0(\mathfrak{A}_{2p-3}^*) = N(2^2, p, 2p - 3)$ and $0(\mathfrak{A}_{2p-1}^*) = N(2^2, p, 2p - 1) - 4p$. The value of $N(N, q, m)$ for $m < 2q$ is given by Lem. XI. Computing with these facts we arrive at the equality stated in the present lemma.

Proof of Lemma XIX. According to Lem. XV we have $\mathfrak{G}_1(p - 2, 1)^* = \mathfrak{U}^2(p - 2, 1)^*$ which has the character $\{\lambda^2(p - 2, 1)\} = \{p, p - 1\} = \{1\}D^{p-1} = \{1\}$. This establishes part (i).

Part (ii) follows from application of Lem. XV and XVI for $(\mu) = (p - 3)$.

According to Lem. XIV, XVI, and XVII, $\mathfrak{U}^1(p - 1)^*$ is the only indecomposable constituent of Π_{2p-1}^* which does not appear in $\Pi_{2p-3}^* \times D$. Hence, by Lem. XVI and XVIII the order of the enveloping algebra of

$$\left\| \begin{matrix} \mathfrak{U}^1(p-1)^* & \\ & \Pi_{2p-3}^* \times D \end{matrix} \right\|$$

is just $(p - 1)^2$ more than the order of $\Pi_{2p-3}^* \times D$. Reference to Lem. XVI and study of the characters of irreducible representations appearing there shows that $\mathfrak{G}_1(0)^*$ does not appear twice in any indecomposable constituent of $\Pi_{2p-3}^* \times D$. Hence, the radical constituent $\mathfrak{G}_1^1(0)^*$ of $\mathfrak{U}^1(p - 1)^*$ cannot appear in $\Pi_{2p-3}^* \times D$. Now by Th. VII and XI the degree of $\mathfrak{G}_1(0)^* = \mathfrak{G}(p - 1, 1)^*$ is $p - 1$; and so $\mathfrak{G}_1^1(0)^*$ contains $(p - 1)^2$ independent matrices. As a consequence of this, every remaining constituent of $\mathfrak{U}^1(0)^*$ must appear somewhere in $\Pi_{2p-3}^* \times D$. We have seen in Parts (i) and (ii) of the present lemma that the irreducible constituents of $\mathfrak{U}^1(0)^*$ occur also in $\mathfrak{U}^1(p - 2, 1)^*$; and a study of the characters shows that $\mathfrak{G}_1(0)^*$ occurs nowhere else in Π_{2p-1}^*. Now $\mathfrak{G}_0^1(0)^*$ can only be equal to some constituent appearing (in $\Pi_{2p-3}^* \times D$) below $\mathfrak{G}_1(0)^*$ and to the left of $\mathfrak{G}(1)^*$; this forces Part (iii) of our lemma. The same considerations relative to $\mathfrak{G}_1^0(0)^*$ gives Part (iv). This completes the proof of Lemma XIX.

University of Michigan

Bibliography

1. R. Brauer and C. Nesbitt, *On the Modular Characters of Groups*, Annals of Math. (2), vol. 42 (1941), pp. 556–590.
2. R. Bauer, *Investigations on Group Characters*, Annals of Math. (4), vol. 42 (1941), pp. 936–958.
3. D. E. Littlewood, *The Theory of Group Characters*, Oxford, 1940.
4. F. D. Murnaghan, *The Theory of Group Representations*, Baltimore, 1938.
5. T. Nakayama, *Some Studies on Regular Representations, Induced Representations, and Modular Representations*, Annals of Math. (2), vol. 39 (1938), pp. 361–369.
6. T. Nakayama, *On Some Modular Properties of Irreducible Representations of a Symmetric Group I*, Japanese Journal of Math., vol. XVII (1940), pp. 89–108; II, Japanese Journal of Math., vol. XVII (1941), pp. 411–423.
7. R. M. Thrall and C. J. Nesbitt, *On the Modular Representations of the Symmetric Group*, Annals of Math. (4), vol. 43 (1942), pp. 656–670.
8. R. M. Thrall, *On the Decomopsition of Modular Tensors* I, Annals of Math. (4), vol. 43 (1942), pp. 671–684.
9. H. Weyl, *Algebraic Theory of Numbers*, Princeton, 1940.
10. H. Weyl. *The Classical Groups*, Princeton, 1939.

Annals of Mathematics
Vol. 45, No. 4, October, 1944

CONSTRUCTION OF CENTRAL SIMPLE ASSOCIATIVE ALGEBRAS

By N. Jacobson

(Received December 20, 1943)

It is well known that any central division algebra $\mathfrak{A}$ over Φ contains a subfield P of dimensionality n if n^2 is the dimensionality of $\mathfrak{A}$ over Φ. Moreover the field P may be chosen to be separable over Φ and there exist algebras of the form $\mathfrak{B} = \mathfrak{A} \times \Phi_r$ that contain normal and separable subfields of dimensionality nr over Φ. The algebra $\mathfrak{B}$ is a normal crossed product and the theory of these algebras has proved to be of great importance in the study of central division algebras. Nevertheless the passage from $\mathfrak{A}$ to $\mathfrak{B}$ is somewhat artificial. A more direct procedure would be to study $\mathfrak{A}$ relative to any one of its maximal subfields P or, at any rate, relative to a maximal separable subfield. This is true especially since it is not known whether or not $\mathfrak{A}$ possesses a normal separable subfield of dimensionality n. The beginning of the study of an algebra relative to an arbitrary subfield was made by Professor Wedderburn in a beautiful paper appearing in 1921.[1] Here Wedderburn showed that $\mathfrak{A}$ can be regarded as a vector space over P and that one obtains in this way a representation of P by matrices with elements in P.

In a recent paper[2] the present author has considered the general theory of self-representations of a field. As we show here, this theory can be applied to investigate the structure of a central simple algebra of dimensionality n^2 relative to any separable subfield of dimensionality n. We shall show that the self-representations of $\mathfrak{A}$ defined by P are the regular self-representations and that $\mathfrak{A}$ may be written as a type of crossed product (P, E, σ) where E denotes a regular self-representation and σ is a factor set. We obtain associativity conditions, conditions that $\mathfrak{A}$ be simple and that $\mathfrak{A}$ be the complete matrix algebra Φ_n. As an application of the theorem that the automorphisms of Φ_n are all inner we obtain necessary and sufficient conditions that an element of P_n have the form $\alpha^E\alpha^{-1}$, α in P and $\alpha \to \alpha^E$ a regular self-representation.

1. Self-representations of fields

In this section we collect some of the results of the theory of self-representations that we require in the sequel.[3] We recall first the definition of a double P-module $\mathfrak{A}$, P a field, to be a left and a right P-module satisfying the following conditions:

1. $$1x = x = x1$$
2. $$(\alpha x)\beta = \alpha(x\beta)$$
3. $$(\mathfrak{A}: P_R) = n < \infty.$$

[1] *On division algebras*, Trans. Am. Math. Soc. **22** (1921), 129–135.

[2] *An extension of Galois theory to non-normal and non-separable fields*, Am. Jour. of Math. vol. 66 (1944), pp. 1–29 referred to hereafter as E.

[3] Cf. E for the results stated in this section.

Here P_R denotes the set of right transformations $x \to x\alpha$. If we choose a right basis $x_1, \cdots, x_n$ for $\mathfrak{A}$, we may write $\alpha x_i = \sum x_\lambda(\alpha E_{\lambda i})$. Then the correspondence $\alpha \to \alpha^E = (\alpha E_{\lambda i})$ is a self-representation of P, i.e. a representation of P by matrices with elements in P. Conversely any self-representation may be obtained in this way. A change of basis from $x_1, \cdots, x_n$ to $y_1, \cdots, y_n$ where $y_i = \sum x_\lambda \mu_{\lambda i}$ replaces the self-representation E by the self-representation F where $\alpha^F = M^{-1}\alpha^E M$, $M = (\mu_{ij})$. We call F and E similar.

If $\alpha \to \alpha^E$ is a self-representation, then $\alpha \to \alpha^{E'} \equiv (\alpha^E)'$ the transposed of α^E is also a self-representation. Let Φ be the subfield of P of elements γ that are fixed in the sense that γ^E is the diagonal matrix $\{\gamma, \cdots, \gamma\}$ and suppose that P is finite and separable over Φ. Then $P = \Phi(\theta)$ and the self-representations E and E' are completely determined by the matrix θ^E. Since θ^E is similar to $\theta^{E'}$ the self-representations E and E' are similar.

We recall also the definition of a composite (K, S, T) of P with itself as a system consisting of a commutative ring K and two isomorphisms S and T between P and subfields P^S and P^T of K such that

$$1)\ 1^S = 1^T, \qquad 2)\ K = P^S P^T, \qquad 3)\ (K: P^T) \text{ is finite.}$$

If $\mathfrak{A}$ is the module associated with the self-representation E, then $\mathfrak{A}$ determines a composite $(P_L P_R, L, R)$ where L is the correspondence between α and the left multiplication $x \to \alpha x$ and R is the correspondence between α and $x \to x\alpha$. On the other hand if (K, S, T) is a composite, K is a double P-module module relative to the multiplications $\alpha x \equiv \alpha^S x = x\alpha^S$ and $x\alpha \equiv \alpha^T x = x\alpha^T$.

There is also a second way in which a composite (K, S, T) determines a double P-module. We choose a basis $\alpha_1^S, \cdots, \alpha_n^S$ for K over P^T. Then for any α in P we may write

$$\alpha^S = \alpha_1^S \mu_1(\alpha)^T + \cdots + \alpha_q^S \mu_q(\alpha)^T \tag{1}$$

and the mappings $M_i : \alpha \to \mu_i(\alpha)$ are single valued. These transformations are endomorphisms in the additive group of P and it is readily seen that the totality of endomorphisms $\sum M_i \bar{\rho}_i$ where $\bar{\rho}$ is the multiplication $\alpha \to \alpha\rho$ in P is an invariant of (K, S, T), i.e. the sets of this type determined by any pair of bases $\alpha_1^S, \cdots, \alpha_n^S$ and $\beta_1^S, \cdots, \beta_n^S$ are identical. We call the set of endomorphisms $\sum M_i \bar{\rho}_i$ the relations space $\mathfrak{S}$ of the composite.

Suppose that the multiplication table of the α_i^S is

$$\alpha_i^S \alpha_j^S = \sum \alpha_k^S \epsilon_{ij}^T. \tag{2}$$

Then a simple computation shows that

$$\bar{\rho} M_i = \sum M_j \bar{\epsilon}_{ij\lambda}(\bar{\rho} M_\lambda). \tag{3}$$

Thus $\mathfrak{S}$ is also invariant under left multiplication by the elements $\bar{\rho}$. It can be shown that $\mathfrak{S}$ has dimensionality n over P (on the right). Hence if we set $\rho M \equiv \bar{\rho} M$ and $M\rho \equiv M\bar{\rho}$ for M in $\mathfrak{S}$, we may consider $\mathfrak{S}$ as a double P-module. The matrix corresponding to ρ in the self-representation determined by the basis

$M_1, \cdots, M_n$ of $\mathfrak{S}$ is because of (3), $\rho^G = (\rho G_{ij})$ where $\rho G_{ji} = \sum \epsilon_{ij\lambda}(\rho M_\lambda)$. On the other hand, if we regard K as a double P-module as before and we use the right basis $\alpha_1^S, \cdots, \alpha_n^S$ for P we have for any ρ in P

$$\begin{aligned}\rho\alpha_j^S \equiv \rho^S\alpha_j^S &= \sum \alpha_\lambda^S\mu_\lambda(\rho)^T\alpha_j^S \\ &= \sum \alpha_i^S\epsilon_{i\lambda j}^T(\rho M_\lambda)^T.\end{aligned}$$

Thus the self-representation determined by this basis is $\rho^E = (\rho E_{ij})$ where $\rho E_{ij} = \sum \epsilon_{i\lambda j}(\rho M_\lambda)$. Since P is commutative $\epsilon_{i\lambda j} = \epsilon_{ij\lambda}$. Hence $G_{ji} = E_{ij}$.

THEOREM 1. *The self-representations determined by the relations space of a composite* (K, S, T) *are the transposed representations of the representations determined by* K.

2. A special generation of Φ_n

A composite of particular interest is obtained by forming the direct product K over Φ of P by itself. Here Φ is a subfield of P such that $(\mathrm{P}:\Phi) = n < \infty$ and $K = \mathrm{P}^S \times \mathrm{P}^T$ where $\gamma^S = \gamma^T$ if $\gamma \in \Phi$ and $(K:\mathrm{P}^T) = (K:\mathrm{P}^S) = (\mathrm{P}:\Phi)$. If $\alpha_1, \cdots, \alpha_n$ is a basis for P over Φ then $\alpha_1^S, \cdots, \alpha_n^S$ is a basis for K over P^T. Thus if $\alpha^S = \sum \alpha_i^S\mu_i(\alpha)^T$ then the elements $\mu_i(\alpha)$ are in Φ. The representations determined by K are called *regular representations* and those determined by the bases $\alpha_1^S, \cdots, \alpha_n^S$ in which $\alpha_1^S = 1^S$ are the *special* regular representations of P. The matrices of a special regular representation belong to Φ_n.

A characteristic property of the direct product is that its relations space $\mathfrak{S}$ is a ring.[4] Moreover it can be shown that $\mathfrak{S}$ is the complete set of endomorphisms in P commutative with the endomorphisms $\bar{\gamma}$, γ in Φ. It follows that $\mathfrak{S} \cong \Phi_n$.

Now let $\alpha_1, \cdots, \alpha_n$ be a basis of P such that $\alpha_1 = 1$. Then for any γ in Φ, $\gamma^S = 1^S\gamma^T = \alpha_1^S\gamma^T$ and hence $\gamma M_1 = \gamma$ and $\gamma M_j = 0$ if $j \neq 1$. Since $\alpha M_i \in \Phi$ for any α it follows that,

$$M_iM_1 = M_i \qquad M_iM_j = 0 \text{ if } j \neq 1. \tag{4}$$

As before for any ρ in P we have $\bar{\rho}M_i = \sum M_j(\overline{\rho E_{ij}})$ where $\rho \to \rho^E$ is the special regular representation determined by the basis $\alpha_1^S, \cdots, \alpha_n^S$ of K. By means of the isomorphism $\mathfrak{S} \cong \Phi_n$ this proves

THEOREM 2. *Let* P *be an arbitrary field over* Φ *such that* $(\mathrm{P}:\Phi) = n$ *and let* $\rho \to \rho^E = (\rho E_{ij})$ *be a special regular representation of* P *over* Φ. *Then* P *may be regarded as a subfield of* Φ_n *and there exist elements* $y_1, \cdots, y_n$ *in* Φ_n *such that every element in* Φ_n *may be written in one and only one way in the form* $\sum y_i\rho_i$, ρ_i *in* P *where the* y_i *satisfy*

$$\rho y_i = \sum y_j(\rho E_{ij}), \qquad y_iy_j = \delta_{1j}y_i. \tag{5}$$

3. Maximal subfields of a central simple algebra

We suppose now that $\mathfrak{A}$ is any central simple algebra of dimensionality n^2 over Φ and that $\mathfrak{A}$ contains a subfield P of dimensionality n. The latter as-

[4] E, pp. 19–21.

sumption is always satisfied when $\mathfrak{A}$ is a division algebra. We turn $\mathfrak{A}$ into a double P-module by defining αx and $x\alpha$ to be the ordinary products of α in P by x in $\mathfrak{A}$. Since $(\mathrm{P}:\Phi) = n$, $(\mathfrak{A}:\mathrm{P}_R) = n = (\mathfrak{A}:\mathrm{P}_L)$. If a is any element in $\mathfrak{A}$ we denote the left multiplication $x \to ax$ by a_l and the right multiplication $x \to xa$ by a_r. It is known that if $a_1, \cdots, a_s$ are elements of $\mathfrak{A}$ that are independent over Φ then $\sum a_{il}b_{ir} = 0$ only if all the $b_i = 0$. In particular if $\alpha_1, \cdots, \alpha_n$ form a basis for P over Φ then $\sum \alpha_{il}\beta_{ir} = 0$ only if all the $\beta_i = 0$. This implies that the composite $(\mathrm{P}_L\mathrm{P}_R, L, R)$ is equivalent to $(\mathrm{P}^S \times \mathrm{P}^T, S, T)$[5].

Suppose now that P is separable over Φ. Then the regular self-representation of P over Φ is completely reducible into dissimilar irreducible components. Since the self-representation determined by $\mathfrak{A}$ has a composite equivalent to the direct product, it is completely reducible into irreducible components that include all of the components of the regular representation. A comparison of the ranks of these two self-representations shows that they are similar.

THEOREM 3. *Let $\mathfrak{A}$ be a central simple algebra of dimensionality n^2 over Φ and let* P *be a separable subfield of dimensionality n. Then the self-representations of* P *obtained by regarding $\mathfrak{A}$ as a double* P*-module are similar to the regular self-representations of* P *over Φ.*

Now let $x_1, \cdots, x_n$ be a right basis of $\mathfrak{A}$ over P. Then for every α in P we have $\alpha x_i = \sum x_j(\alpha E_{ji})$, where the self-representation $\alpha \to \alpha^E$ is regular. Moreover, if

$$x_ix_j = \sum x_\mu\sigma_{\mu ij} \tag{6}$$

then the product of any two elements of $\mathfrak{A}$ is completely determined by the multiplication of elements in P, by the self-representation E and by the *factor set* $\sigma = \{\sigma_{ijk}\}$. Since $(x_ix_j)x_k = x_i(x_jx_k)$

$$\sum_{\lambda,\mu} \sigma_{\rho\lambda\mu}(\sigma_{\lambda ij}E_{\mu k}) = \sum_{\lambda} \sigma_{\rho i\lambda}\sigma_{\lambda jk} \tag{7}$$

and since $\rho(x_ix_j) = (\rho x_i)x_j$ for all ρ in P

$$\sum_{\lambda,\mu} \sigma_{\nu\lambda\mu}(\rho E_{\lambda i}E_{\mu j}) = \sum (\rho E_{\nu\lambda})\sigma_{\lambda ij}. \tag{8}$$

4. Crossed products

We suppose now that P is any finite separable extension of Φ of dimensionality n and let $\mathfrak{A}$ be a double P-module whose representations are regular. Let $x_1, \cdots, x_n$ be a right P-basis for $\mathfrak{A}$ and for $x = \sum x_i\rho_i$, $x' = \sum x_j\rho_j'$ define

$$xx' = \sum x_\mu\sigma_{\mu i\lambda}(\rho_iE_{\lambda j})\rho_j' \tag{9}$$

where the elements σ_{ijk} are arbitrary elements in P that satisfy (7) and (8). We wish to show that $\mathfrak{A}$ is an associative algebra over Φ. We note first that by (9)

[5] Equivalence of composites is defined as for field composites. Cf. E, p. 5.

$$(10) \qquad (xx')\rho = x(x'\rho).$$

Moreover $(\sum x_i\rho_i)(\sum x_j\rho_j') = \sum x_\mu \sigma_{\mu i \lambda}(\rho_i E_{\lambda j})\rho_j' = \sum (x_i x_\lambda)\rho_i E_{\lambda j}\rho_j'$. Since $\rho \rightarrow \rho^E$ is a representation this implies that

$$(11) \qquad (x\rho)x' = x(\rho x').$$

We note also that $\rho(x_i x_j) = (\rho x_i)x_j$ and using these relations we may prove that

$$(x_i\rho)((x_j\sigma)x_k\tau) = \sum_{\lambda,\mu,\nu} x_i(x_\mu x_\nu)(\rho E_{\mu j} E_{\nu\lambda}\sigma E_{\lambda k}\tau)$$

and

$$((x_i\rho)(x_j\sigma))x_k\tau = \sum_{\lambda,\mu,\nu} (x_i x_\mu)x_\nu(\rho E_{\mu j} E_{\nu\lambda}\sigma E_{\lambda k}\tau).$$

Since by (7) $x_i(x_\mu x_\nu) = (x_i x_\mu)x_\nu$ this proves that $\mathfrak{A}$ is associative. If $\gamma \in \Phi$, $\gamma x = x\gamma$ and hence

$$(xx')\gamma = x(x'\gamma) = (x\gamma)x'$$

so that $\mathfrak{A}$ is an algebra over Φ. Evidently $(\mathfrak{A}\colon \Phi) = (\mathfrak{A}\colon \mathrm{P})(\mathrm{P}\colon \Phi) = n^2$.

THEOREM 4. *Let* P *be a separable extension of dimensionality n over* Φ, $\alpha \rightarrow \alpha^E$ *a regular self-representation of* P *over* Φ *and* $\{\sigma_{ijk}\}$ *a set of elements in* P *satisfying* (7) *and* (8). *Then the totality of elements* $\sum x_i\rho_i$ *of the n-dimensional space over* P *is an algebra if multiplication is defined by* (9).

We shall call $\mathfrak{A}$ the *crossed product* of P with regular self-representation E and factor set $\sigma = \{\sigma_{ijk}\}$ and we write $\mathfrak{A} = (\mathrm{P}, E, \sigma)$. The conditions (7) and (8) will be called the *associativity conditions* on the factor set σ. By the preceding section we have

THEOREM 5. *Let* $\mathfrak{A}$ *be a central simple algebra of dimensionality* n^2 *over* Φ *and let* P *be a separable subfield of* $\mathfrak{A}$ *of dimensionality n. Then* $\mathfrak{A}$ *is a crossed product* (P, E, σ).

The associativity conditions (8) take on a simpler form if E is a special regular representation. For in this case ρE_{ij} are in Φ and hence $\rho E_{ij} E_{kl} = \delta_{kl}(\rho E_{ij})$. The equations (8) reduce to

$$(12) \qquad \sum \sigma_{\nu\lambda j}(\rho E_{\lambda i}) = \sum (\rho E_{\nu\lambda})\sigma_{\lambda ij}.$$

If we set $\sigma_j = (\sigma_{\nu\lambda j})$ this is equivalent to

$$(13) \qquad \sigma_j\rho^E = \rho^E\sigma_j.$$

LEMMA. *If A is a matrix in* P_n *that commutes with all the matrices* ρ^E, *A is a linear combination of the matrices* ρ^E *with coefficients in* P.

A matrix A in P_n commutative with all the ρ^E corresponds to an endomorphism $\bar{A}$ in the direct product $K = \mathrm{P}^S \times \mathrm{P}^T$ that commutes with all the multiplications $x \rightarrow x\rho^S$ and all of the multiplications $x \rightarrow x\rho^T$ and hence with all of the multiplications $x \rightarrow xa$ for arbitrary a in K. Since K is commutative, $\bar{A}$ is a multiplication. It follows that A is a linear combination of the matrices ρ^E using coefficients in P.

Thus the elements σ_j that satisfy (13) are linear combinations of matrices ρ^I. Each σ_j will therefore depend on n parameters in P and hence the factor set σ depends on n^2 parameters.

We suppose again the E is an arbitrary regular representation. Then we may verify that

$$\rho(xx') = (\rho x)x'. \tag{14}$$

In particular if $y_1, \cdots, y_n$ is a second right P-basis for $\mathfrak{A}$ then $\rho(y_i y_j) = (\rho y_i)y_j$. Hence if $y_i y_j = \sum y_\lambda \tau_{\lambda ij}$ and $\rho y_i = \sum y_j(\rho F_{ji})$ then the τ's and the F_{ij} satisfy equations analogous to (7) and (8). If $y_i = \sum x_\lambda \mu_{\lambda i}$ we know that $\alpha^F = M^{-1}\alpha^E M$, $M = (\mu_{ij})$. We may verify also that

$$\tau_{kij} = \sum \bar{\mu}_{k\rho}\sigma_{\rho\lambda\mu}(\mu_{\lambda i}E_{\mu\nu})\mu_{\nu j} \tag{15}$$

where $(\bar{\mu}_{ij}) = M^{-1}$.

THEOREM 6. *If M is any non-singular matrix in P_n, then the crossed products* (P, E, σ) *and* (P, F, τ) *where* $\alpha^F = M^{-1}\alpha^E M$ *and* τ *satisfies* (15) *are isomorphic.*

5. Conditions for simplicity

We consider now the conditions that $\mathfrak{A} = (P, E, \sigma)$ be central and simple. For this purpose we investigate the structure of the double P-module, $\mathfrak{A}$. By definition $\mathfrak{A}$ is isomorphic to the double P-module $K = P^S \times P^T$ obtained from the direct product (K, S, T). Hence we consider K. The submodules of K are the (two-sided) ideals of the ring K. Since P is separable over Φ, $K = K_1 \oplus K_2 \oplus \cdots \oplus K_s$ where the K_i are fields. The self-representations determined by the K_i are irreducible and dissimilar and any irreducible self-representation of P over Φ is similar to one of these. In particular one of the K_i, say K_1, is associated with the identity self-representation $\alpha \to \alpha$. Any submodule of K is a direct sum of certain of the K_i.

If $\alpha \to \alpha^{E_1}$ and $\alpha \to \alpha^{E_2}$ are self-representations of P the product $E_1 \times E_2$ is the self-representation obtained by replacing the elements αE_{ij1} of α^{E_1} by the matrices $\alpha E_{ij1}^{E_2}$ that are associated with these elements in E_2.[6] Let $\mathfrak{A}_1$ be the double P-module associated with E_1 and $x_1, \cdots, x_m$ a basis such that $\alpha x_i = \sum x_\lambda(\alpha E_{\lambda i1})$ and similarly let $\mathfrak{A}_2$ be a double P-module with basis $y_1, \cdots, y_r$ such that $\alpha y_j = \sum y_\mu(\alpha E_{\mu j2})$. Then it is possible to define a double P-module $\mathfrak{A}_1 \times \mathfrak{A}_2$ having a right basis $x_i \times y_j$ such that $\alpha(x_i \times y_j) = \sum x_\lambda \times y_\mu(\alpha E_{\lambda i1}E_{\mu j2})$.[7] This product module is independent of the choice of the bases in $\mathfrak{A}_i$. It follows from this that if F_1 is a self-representation similar to E_1 and F_2 is similar to E_2 then $F_1 \times F_2$ is similar to $E_1 \times E_2$.

Now let E_i be an irreducible self-representation associated with K_i and E_j one associated with K_j and form $E_i \times E_j$. This representation is completely

---

[6] Cf. E. p. 12 or L. KALUJNINE, *Sur la theorie de Galois de Corps non-galoisiennes separables*, Comptes Rendus de l'Acad. des Sciences, *214* (1942), 597–599.

[7] E, p. 12.

reducible into irreducible representations determined by, say, $K_{i_1}, \cdots, K_{i_r}$. If we replace E_i and E_j by F_i and F_j two other self-representations associated with K_i and K_j respectively, we may associate with $F_i \times F_j$ the same set $K_{i_1}, \cdots, K_{i_r}$. Hence we may define the set $(K_{i_1}, \cdots, K_{i_r})$ to be the product $K_i \times K_j$ of K_i and K_j. It is known that relative to this product the set of K_i forms a hypergroup $\mathfrak{H}$: the product is associative, K_i acts as an identity and for each K_i there exists a $K_{i'}$ such that $K_i \times K_{i'} = (K_1, \cdots)$ and an element $K_{i''}$ such that $K_{i''} \times K_i = (K_1, \cdots)$.[8]

We now apply these results to $\mathfrak{A} = (\mathrm{P}, E, \sigma)$. Then we see that $\mathfrak{A} = \mathfrak{A}_1 \oplus \mathfrak{A}_2 \oplus \cdots \oplus \mathfrak{A}_s$ where $\mathfrak{A}_i$ is the double P-module corresponding to K_i. Any submodule of $\mathfrak{A}$ is a direct sum of certain of the $\mathfrak{A}_i$. We now define the *product* $\mathfrak{A}_i\mathfrak{A}_j$ to be the set of sums $\sum a_i a_j$ where $a_i \in \mathfrak{A}_i$ and $a_j \in \mathfrak{A}_j$. By (10) and (14) $\mathfrak{A}_i\mathfrak{A}_j$ is a submodule of $\mathfrak{A}$. Thus $\mathfrak{A}_i\mathfrak{A}_j$ is a direct sum of certain of the $\mathfrak{A}_i$. Let $x_1, \cdots, x_m$ be a right basis for $\mathfrak{A}_i$ and $y_1, \cdots, y_r$ a right basis for $\mathfrak{A}_j$. Any element of $\mathfrak{A}_i\mathfrak{A}_j$ is a sum of elements of the form $(\sum x_i\xi_i)(\sum y_j\eta_j)$ and this may be written as $\sum x_i y_j \zeta_{ij}$ if we use (11). Thus the right module generated by the elements $x_i y_j$ is the whole module $\mathfrak{A}_i\mathfrak{A}_j$. We consider now the module $\mathfrak{A}_i \times \mathfrak{A}_j$. It is readily verified that the mapping $\sum (x_i \times y_j)\zeta_{ij} \rightarrow \sum x_i y_j \zeta_{ij}$ is a homomorphism between the P-modules $\mathfrak{A}_i \times \mathfrak{A}_j$ and $\mathfrak{A}_i\mathfrak{A}_j$. Hence the irreducible components of $\mathfrak{A}_i\mathfrak{A}_j$ are contained among the irreducible components of $\mathfrak{A}_i \times \mathfrak{A}_j$. It follows from our definition of $K_i \times K_j$ that if $K_i \times K_j = (K_{i_1}, \cdots, K_{i_r})$, then $\mathfrak{A}_i\mathfrak{A}_j = \mathfrak{A}_{i_1} \oplus \cdots \oplus \mathfrak{A}_{i_t}$, $t \leqq r$ if the K's are suitably ordered. Since $K_1 \times K_j = K_j = K_j \times K_1$, $\mathfrak{A}_1\mathfrak{A}_j = 0$ or $\mathfrak{A}_j$ and $\mathfrak{A}_j\mathfrak{A}_1 = 0$ or $\mathfrak{A}_j$.

We assume now that $\mathfrak{A}_j\mathfrak{A}_1 = \mathfrak{A}_j = \mathfrak{A}_1\mathfrak{A}_j$ for all $\mathfrak{A}_j$. In particular $\mathfrak{A}_1^2 = \mathfrak{A}_1$. Hence if u is an element $\neq 0$ in $\mathfrak{A}_1$, $u^2 = u\delta$ where $\delta \neq 0$. Then $e = u\delta^{-1}$ is idempotent. Since $\mathfrak{A}_1\mathfrak{A} = \mathfrak{A} = \mathfrak{A}\mathfrak{A}_1$, $e\mathfrak{A} = \mathfrak{A} = \mathfrak{A}e$. Hence e is the identity 1 of $\mathfrak{A}$. Now $\mathfrak{A}$ contains the subfield $\mathfrak{A}_1 = 1\mathrm{P}$ isomorphic to P. Since $\alpha x = \alpha(1x) = (\alpha 1)x = (1\alpha)x$ and $x\alpha = (x1)\alpha = x(1\alpha)$ the module operations in $\mathfrak{A}$ are the left and the right multiplications by the elements of $1\mathrm{P}$. If $\mathfrak{B}$ is a two-sided ideal, $\mathfrak{B}$ is a submodule of $\mathfrak{A}$ and hence $\mathfrak{B}$ is a sum of certain of the $\mathfrak{A}_i$. This proves

THEOREM 7. *If $\mathfrak{A}$ has an identity there are only a finite number of two-sided ideals in $\mathfrak{A}$. Any two-sided ideal is a sum of certain of the submodules $\mathfrak{A}_i$.*

If we choose a basis $u_1, \cdots, u_n$ of $\mathfrak{A}$ composed of bases for the $\mathfrak{A}_i$, a condition that $\mathfrak{A}$ have an identity may be obtained on the elements μ_{kij} such that $u_i u_j = \sum u_k \mu_{kij}$. This is that the matrices (μ_{ki1}) and (μ_{k1j}) be non-singular. For this implies that $\mathfrak{A}_i\mathfrak{A}_1 \neq 0$ and $\mathfrak{A}_1\mathfrak{A}_i \neq 0$.

We note next that the only elements of $\mathfrak{A}$ that commute with all the elements of $1\mathrm{P}$ are the elements in this field. For if u is such an element then $\alpha u = u\alpha$ for all α. Hence $u \in \mathfrak{A}_1 = 1\mathrm{P}$. Since the only elements of P that are fixed under any regular self-representation are the elements of Φ[9], the only elements

[8] E, p. 25 or KALUJNINE loc. cit. in 6.

[9] E, p. 21.

of $1P$ that commute with all the elements x_i of a P-basis of $\mathfrak{A}$ are the elements in 1Φ. It follows that 1Φ is the center of $\mathfrak{A}$.

THEOREM 8. *If $\mathfrak{A}$ has an identity, $\mathfrak{A}$ is a central algebra.*

We assume next that for each module $\mathfrak{A}_i$ there exists an $\mathfrak{A}_j$ such $\mathfrak{A}_j\mathfrak{A}_i$ contains $\mathfrak{A}_1$. Let $\mathfrak{B}$ be a two-sided ideal $\neq 0$ in $\mathfrak{A}$. Then $\mathfrak{B}$ contains one of the $\mathfrak{A}_i$ and hence $\mathfrak{B}$ contains $\mathfrak{A}_j\mathfrak{A}_i$. It follows that $1 \in \mathfrak{B}$ and $\mathfrak{B} = \mathfrak{A}$.

THEOREM 9. *The algebra $\mathfrak{A}$ is simple if* 1) $\mathfrak{A}_1\mathfrak{A}_i = \mathfrak{A}_i = \mathfrak{A}_i\mathfrak{A}_1$ *for all $\mathfrak{A}_i$ and* 2) *for each $\mathfrak{A}_i$ there exists an $\mathfrak{A}_j$ such that $\mathfrak{A}_j\mathfrak{A}_i$ contains $\mathfrak{A}_1$.*

Conversely suppose that $\mathfrak{A}$ is simple. Then $\mathfrak{A}$ contains an identity. Since $a1 = 1\alpha$, $1 \in \mathfrak{A}_1$. Hence $\mathfrak{A}_1\mathfrak{A}_i = \mathfrak{A}_i = \mathfrak{A}_i\mathfrak{A}_1$.

We consider now the set $\mathfrak{M} = \mathfrak{A}_L P_R = P_R \mathfrak{A}_L$ of endomorphisms of the form $a_{1l}\rho_{1r} + \cdots + a_{sl}\rho_{sr}$ where a_l is a left multiplication $x \to ax$, a in $\mathfrak{A}$ and ρ_r is a right multiplication by an element of P. We recall that if $a_1, \cdots, a_s$ are linearly independent over Φ then $a_{1l}\rho_{1r} + \cdots + a_{sl}\rho_{sr} \neq 0$ unless all the ρ_i are 0. Thus $(\mathfrak{M}: P_R) = (\mathfrak{A}: \Phi) = n^2$. On the other hand the elements of $\mathfrak{M}$ are endomorphisms in $\mathfrak{A}$ that commute with all the elements of P_R. Since the dimensionality of $\mathfrak{A}$ over P_R is n, it follows that $\mathfrak{M}$ is the complete set of endomorphisms in $\mathfrak{A}$ commutative with the elements of P_R. This implies that $\mathfrak{M}$ is an irreducible set of endomorphisms.

Now consider any one of the $\mathfrak{A}_i$. Form $\mathfrak{A}\mathfrak{A}_i$. This is a submodule of $\mathfrak{A}$ and a left ideal. Hence $\mathfrak{A}\mathfrak{A}_i$ is a subgroup $\neq 0$ of $\mathfrak{A}$ which is invariant under all the endomorphisms in $\mathfrak{M}$. Since $\mathfrak{M}$ is irreducible $\mathfrak{A}\mathfrak{A}_i = \mathfrak{A}$. It follows that there exists an $\mathfrak{A}_j$ such that $\mathfrak{A}_j\mathfrak{A}_i$ contains $\mathfrak{A}_1$. Similarly there exists an $\mathfrak{A}_k$ such that $\mathfrak{A}_i\mathfrak{A}_k$ contains $\mathfrak{A}_1$. This proves

THEOREM 10. *If $\mathfrak{A}$ is a simple algebra, the irreducible modules $\mathfrak{A}_i$ form a hypergroup H under multiplication.*

We have also noted above the relation between H and the hypergroup $\mathfrak{H}$ of the K_i, namely, there is a (1–1) correspondence $K_i \to \mathfrak{A}_i$ between $\mathfrak{H}$ and H such that if $K_i \times K_j = (K_{i_1}, \cdots, K_{i_r})$ then $\mathfrak{A}_i\mathfrak{A}_j = (\mathfrak{A}_{i_1}, \cdots, \mathfrak{A}_{i_t})$ where $t \leqq r$.

We consider now the special case where P is normal. Here the irreducible self-representations of P are of rank 1 and hence they are identical with the automorphism $\alpha \to \alpha^S$ of the Galois group $\mathfrak{G}$ of P over Φ. The modules $\mathfrak{A}_i$ are one dimensional over P and are in (1–1) correspondence with the elements S. We may denote the correspondent of S by $\mathfrak{A}_S$. Then the condition that $\mathfrak{A}$ be simple is that $\mathfrak{A}_S\mathfrak{A}_T = \mathfrak{A}_{ST}$. We choose any element $u_S \neq 0$ in $\mathfrak{A}_S$. Then the condition that $\mathfrak{A}$ be simple is that $u_S u_T = u_{ST}\mu_{S,T}$ where $\mu_{S,T} \neq 0$. These conditions for normal crossed products are well-known, as are also the associativity conditions on the $\mu_{S,T}$ that one obtains from (7).

6. Conditions that $\mathfrak{A} \cong \Phi_n$

We have seen that a change of basis from $x_1, \cdots, x_n$ to $y_1, \cdots, y_n$ in a crossed product $\mathfrak{A}$ replaces the self-representation E associated with the x's by a similar self-representation $\alpha \to \alpha^F = M^{-1}\alpha^E M$ and replaces the x-factor set σ by a factor

set τ defined by (15). If M is a matrix commutative with all the α^{E} then $F = E$. In this case we shall say that the factor sets σ and τ are *associates* ($\sigma \sim \tau$).

In order to determine whether or not a central simple $\mathfrak{A}$ is the complete matrix algebra, it suffices to consider $\mathfrak{A}$ in the form (P, E', σ) where E' is the transposed of a special regular representation. Now we have seen in Theorem 2 that if 1 denotes the factor set $\{\lambda_{ijk}\}$ where $\lambda_{ijk} = \delta_{k1}\delta_{ij}$ then $(\mathrm{P}, E', 1) \cong \Phi_n$ or $(\mathrm{P}, E', 1) \sim 1$. The usual proof making use of the fact that an isomorphism between simple subalgebras of a central simple algebra $\mathfrak{A}$ may be extended to an automorphism in $\mathfrak{A}$ shows that conversely if $(\mathrm{P}, E', \sigma) \sim 1$ then $\sigma \sim 1$.

THEOREM 11. *A necessary and sufficient condition that a central simple algebra* $\mathfrak{A} = (\mathrm{P}, E', \sigma) \sim 1$ *is that* $\sigma \sim 1$.

Suppose now that $\mathfrak{A} = (\mathrm{P}, E', 1)$ and that $y_1, \cdots, y_n$ is a P-basis such that $\rho y_i = \sum y_j(\rho E_{ij})$ and $y_i y_j = \delta_{1j} y_i$. Let $z_1, \cdots, z_n$ be a second right P-basis such that $\rho z_i = \sum z_j(\rho E_{ij})$ and $z_i z_j = \delta_{1j} z_i$. Then $z_i = \sum y_j \mu_{ji}$ where the matrix M satisfies the following two conditions:

$$\alpha^{E'} M = M\alpha^{E'} \quad \text{for all } \alpha, \tag{16}$$

$$\delta_{j1}\, \delta_{ik} = \sum_{\rho,\nu} \bar{\mu}_{k\rho}(\mu_{\rho i} E_{\nu 1})\mu_{\nu j}. \tag{17}$$

For (17) is obtained from (15) by setting $\tau_{kij} = \lambda_{kij}$, $\sigma_{\rho\lambda\mu} = \lambda_{\rho\lambda\mu}$ and by replacing $E_{\mu\nu}$ by $E_{\nu\mu}$. From (17) we obtain

$$(\mu_{\rho i} E_{\nu 1})\mu_{\nu j} = \delta_{j1}\mu_{\rho i} \tag{18}$$

and conversely (18) implies (17). Now let M be any non-singular matrix in P_n satisfying (16) and (18). Then the elements $z_i = \sum y_j \mu_{ji}$ will satisfy the same multiplication table as the y_i. Thus (16) and (18) are the conditions that the mapping $\sum y_i \rho_i \to \sum z_i \rho_i$ be an automorphism A in $(\mathrm{P}, E', 1)$. It is clear that A leaves the elements of P invariant. Hence A is an inner automorphism of the form $x \to \beta x \beta^{-1}$. In particular $z_i = \beta y_i \beta^{-1} = \sum y_j(\beta E_{ij})\beta^{-1} = \sum y_j \mu_{ji}$ and so $M = \beta^{E'}\beta^{-1}$. Conversely if M has this form then (16) and (18) hold.

THEOREM 12. *Let* P *be an extension of* Φ *such that* $(\mathrm{P}: \Phi) = n$ *and let* E *be a special regular representation of* P *over* Φ. *Then the conditions* (16) *and* (18) *are necessary and sufficient that the non-singular matrix* M *in* P_n *have the form* $\beta^{E'}\beta^{-1}$.

We recall that (16) holds if and only if M is a linear combination of matrices $\alpha^{E'}$ using coefficients in P. It may be remarked also that separability is not required in Theorem 11 since P was arbitrary in Theorem 2.

THE JOHNS HOPKINS UNIVERSITY.

Annals of Mathematics
Vol. 45, No. 4, October, 1944

THE AVERAGE MEASURE OF QUADRATIC FORMS WITH GIVEN DETERMINANT AND SIGNATURE

By Carl Ludwig Siegel

(Received May 17, 1944)

1. Let h_d be the number of classes of primitive binary quadratic forms $Q(x, y) = ax^2 + bxy + cy^2$ with positive discriminant $d = b^2 - 4ac$; we use the narrow definition of equivalence: two quadratic forms are said to belong to the same class if they can be transformed into each other by unimodular substitutions with determinant $+1$. Let d be no square number, and define $\epsilon_d = (t + u\sqrt{d})/2$, where t, u are the smallest positive integral solutions of $t^2 - du^2 = 4$.

When $d = 4k$ is divisible by 4, then b is even and $Q(x, y)$ is a properly primitive quadratic form of determinant k, in the notation of Gauss. In his *Disquisitiones Arithmeticae* Gauss[1] stated that the mean value of the expression $h_{4k} \log \epsilon_{4k}$ is asymptotically equal to $\frac{2\pi^2}{7\zeta(3)} k^{\frac{1}{2}}$; this means that

$$(1) \qquad \sum_{k \leqq N} h_{4k} \log \epsilon_{4k} \sim \frac{4\pi^2}{21\zeta(3)} N^{3/2} \qquad (N \to \infty).$$

No proof of this statement has been published until now.

In the first chapter we shall prove the assertion of Gauss in two different ways. Both methods may also be used to determine more generally the average of $h_d \log \epsilon_d$ in any class of residues $d \equiv d_0 \pmod{m}$; however, we shall consider only the cases $m = 4$, $m = 1$, and we shall prove the formula

$$(2) \qquad \sum_{k \leqq N} h_k \log \epsilon_k = \frac{\pi^2}{18\zeta(3)} N^{3/2} + O(N \log N).$$

The corresponding results for the wide definition of equivalence (unimodular substitutions with determinant ± 1) are an immediate consequence, in virtue of the relationship $h \log \epsilon = 2h_0 \log \epsilon_0$, where h_0 denotes the class number in the wide sense, $\epsilon_0 = (t_0 + u_0\sqrt{d})/2$, and t_0, u_0 are the smallest positive integral solutions of $t_0^2 - du_0^2 = \pm 4$.

2. In the second chapter we investigate the analogous problem for quadratic forms $\mathfrak{S}[\mathfrak{x}]$ with m variables and prescribed signature n, $m - n$, the matrix $\mathfrak{S}$ being integral. Let $\mathfrak{T}$ be a variable real symmetric matrix of signature n, $m - n$, which lies in a domain T; let Y denote the domain in the space of all real m-rowed matrices $\mathfrak{Y}$ which is mapped into T by the condition $\mathfrak{S}[\mathfrak{Y}] = \mathfrak{T}$; let Y_0 be a fundamental region in Y with respect to the group of units of $\mathfrak{S}$. Denote by $v(Y_0)$ and $v(T)$ the Euclidean volumes of Y_0 and T, where the m^2 elements of $\mathfrak{Y}$

[1] C. F. Gauss, *Werke*, I (Zweiter Abdruck, Göttingen, 1870); p. 369 and p. 466.

and the $(m(m+1))/2$ independent elements of $\mathfrak{T}$ are taken as rectangular Cartesian coordinates, and let T converge to the single point $\mathfrak{S}$. It is known that

$$\lim_{T\to\mathfrak{S}} v(Y_0)/v(T) = \rho(\mathfrak{S}) \tag{3}$$

exists, except in the trivial case of a rationally decomposable binary form; in this case we define $\rho(\mathfrak{S}) = 0$. Our object is the proof of the following

THEOREM: *Let $\mathfrak{S}$ run through a system of representatives of all classes of given signature $n, m-n$, whose determinants have the absolute value $S \leqq N$; then*

$$\sum_{S\leqq N} S\rho(\mathfrak{S}) \sim \tfrac{1}{2}N \prod_{k=2}^{m} \zeta(k) \qquad (N\to\infty). \tag{4}$$

In the definite case, i.e., for $n = 0$ and $n = m$, this theorem was already given by Minkowski,[2] in a slightly different form. In this case,

$$S^{(m+1)/2}\rho(\mathfrak{S}) = \frac{1}{E(\mathfrak{S})}\prod_{k=1}^{m}\frac{\pi^{k/2}}{\Gamma\left(\frac{k}{2}\right)}, \tag{5}$$

where $E(\mathfrak{S})$ denotes the order of the group of units of $\mathfrak{S}$, and it follows from (4) by partial summation that

$$\sum_{S\leqq N}\frac{2}{E(\mathfrak{S})} \sim \frac{2}{m+1} N^{(m+1)/2}\prod_{k=2}^{m}\left\{\pi^{-k/2}\Gamma\left(\frac{k}{2}\right)\zeta(k)\right\} = \gamma_m N^{(m+1)/2}, \tag{6}$$

say. On the other hand, $E(\mathfrak{S}) = 2$, except when $\mathfrak{S}$ is equivalent to a matrix on the frontier of the reduced domain, and it is easily seen that the sum of the class numbers $H_m(S)$ of all positive quadratic forms of m variables and determinant $S \leqq N$ is asymptotically equal to the left-hand member in (6). Consequently,

$$\sum_{S=1}^{N} H_m(S) \sim \gamma_m N^{(m+1)/2}, \tag{7}$$

and this is Minkowski's formula; *vice versa*, (4) follows from (7) by partial summation. As a matter of fact, Minkowski established a better result than (7), namely the estimate $O(N^{m/2})$ $(m > 2)$, $O(N \log N)$ $(m = 2)$ for the difference of both sides in (7). Since the proof of (4) is decidedly more difficult in the indefinite case $0 < n < m$, we are interested in the simplest appraisal of the error term which is sufficient for the proof of the Theorem, and this leads only to the estimate $o(N)$ for the difference of both sides in (4). It is possible to obtain the estimate $O(N^{\frac{1}{2}})$ $(m > 2)$, $O(N^{\frac{1}{2}} \log N)$ $(m = 2)$ also in the case of an arbitrary signature; however, this is rather complicated for $m \geqq 3$, and we shall derive this more precise result only in case $m = 2$.

[2] H. MINKOWSKI, *Diskontinuitätsbereich für arithmetische Äquivalenz*, Journal für die reine und angewandte Mathematik 129, pp. 220–274 (1905).

I. Binary Forms

3. Let $\chi(k)$ be a non-principal character modulo q and define

$$s_n = \sum_{k=1}^{n} \chi(k) \qquad (n = 1, 2, \cdots).$$

It has been proved by Pólya[3] for proper characters, and by Landau[4] in the general case that

$$|s_n| < cq^{\frac{1}{2}} \log q, \tag{8}$$

where c is an absolute constant. Let N be a positive integer; then

$$\sum_{n=N+1}^{\infty} \chi(n)n^{-1} = \sum_{n=N+1}^{\infty} (s_n - s_{n-1})n^{-1} = \sum_{n=N}^{\infty} s_n \left(\frac{1}{n} - \frac{1}{n+1}\right) - s_N N^{-1}$$

and, by (8),

$$\left| \sum_{n=N+1}^{\infty} \chi(n)n^{-1} \right| < 2cN^{-1}q^{\frac{1}{2}} \log q. \tag{9}$$

In particular, the Legendre-Jacobi-Kronecker symbol $\left(\frac{d}{k}\right)$ is a non-principal character modulo $|d|$, whenever $d \equiv 0$ or 1 (mod 4) and not a square number. It is well known that

$$d^{-\frac{1}{2}} h_d \log \epsilon_d = \sum_{n=1}^{\infty} \left(\frac{d}{n}\right) n^{-1} \qquad (d > 0). \tag{10}$$

We introduce the abbreviations

$$d^{-\frac{1}{2}} h_d \log \epsilon_d = f_d, \qquad \sum_{n=1}^{N} \left(\frac{d}{n}\right) n^{-1} = \sigma_d;$$

it follows from (9) and (10) that

$$|f_d - \sigma_d| < 2cN^{-1}d^{\frac{1}{2}} \log d, \tag{11}$$

for all positive $d \equiv 0$ or 1 (mod 4) which are no squares. On the other hand,

$$|\sigma_d| \leqq \sum_{n=1}^{N} n^{-1} < 1 + \log N, \tag{12}$$

also when d is a square.

---

[3] G. Pólya, *Über die Verteilung der quadratischen Reste und Nichtreste*, Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse, Jahrgang 1918, pp. 21-29.

[4] E. Landau, *Abschätzung von Charaktersummen, Einheiten und Klassenzahlen*, Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse, Jahrgang 1918, pp. 79-97.

Let r be either 0 or 1; let t run through all numbers of the given residue class $t \equiv r \pmod 4$ which lie in the interval $1 \leqq t \leqq N$, and let d run through the set of all those t which are no square numbers. By (11) and (12),

$$\sum_d f_d = \sum_{n=1}^{N} n^{-1} \sum_t \left(\frac{t}{n}\right) + O(N^{\frac{1}{2}} \log N) \qquad (N \to \infty). \tag{13}$$

Define

$$\sum_t \left(\frac{t}{n}\right) = P_r(n) \qquad (n = 1, 2, \cdots, N), \tag{14}$$

and assume first that n is not a square. If n is even, then $P_r(n) = 0$. If n is odd, then $\chi_1(k) = \left(\frac{k}{n}\right)$ is a character modulo n, and not the principal character; consequently, by (8),

$$P_0(n) = \sum_{4k \leqq N} \chi_1(k) = n^{\frac{1}{2}} \log n O(1).$$

Let l denote the highest power of 2 dividing n, then $n/l = s$ is odd and $\left(\frac{t}{n}\right) = \left(\frac{l}{t}\right)\left(\frac{t}{s}\right)$, for odd t. Since $\chi_2(k) = \left(\frac{4l}{k}\right)\left(\frac{k}{s}\right)$ and $\chi_3(k) = \left(\frac{-4l}{k}\right)\left(\frac{k}{s}\right)$ are characters modulo $4n$, both different from the principal character, we obtain, again by (8),

$$P_1(n) = \tfrac{1}{2} \sum_{k \leqq N} (\chi_2(k) + \chi_3(k)) = n^{\frac{1}{2}} \log n O(1).$$

Hence

$$P_r(n) = n^{\frac{1}{2}} \log n O(1), \tag{15}$$

for $r = 0, 1$ and $1 \leqq n \leqq N, n \neq 1, 4, 9, \cdots$.

Next let $n = u^2$ be a square. Then $P_r(n)$ equals the number of integers in the interval $1 \leqq t \leqq N$ which are prime to u and $\equiv r \pmod 4$. Therefore

$$P_r(u_1^2) = \frac{\varphi(u_1)}{4u_1} N + u_1 O(1) \qquad (u_1 \text{ odd}),$$

$$P_r(u_2^2) = \frac{r\varphi(u_2)}{2u_2} N + u_2 O(1) \qquad (u_2 \text{ even}).$$

It follows that

$$\begin{aligned} \sum_{u^2 \leqq N} u^{-2} P_r(u^2) &= \frac{N}{4} \sum_{u_1^2 \leqq N} u_1^{-3} \varphi(u_1) + \frac{rN}{2} \sum_{u_2^2 \leqq N} u_2^{-3} \varphi(u_2) + O(\log N) \\ &= \frac{N}{4} \frac{\zeta(2)}{\zeta(3)} \left(1 + \frac{2r - 1}{7}\right) + O(\sqrt{n}). \end{aligned} \tag{16}$$

By (13), (14), (15), (16),

$$\sum_d f_d = \sum_{n=1}^{N} n^{-1}P_r(n) + O(N^{\frac{1}{2}} \log N)$$

$$(17) \qquad = \frac{\pi^2 N}{24\zeta(3)}\left(1 + \frac{2r-1}{7}\right) + O(N^{\frac{1}{2}} \log N).$$

Applying partial summation, we infer that

$$(18) \qquad \sum_d d^{\frac{1}{2}} f_d = \frac{\pi^2 N^{3/2}}{36\zeta(3)}\left(1 + \frac{2r-1}{7}\right) + O(N \log N),$$

where d runs through all integers in the interval $1 \leqq t \leqq N$ which are $\equiv r \pmod 4$ and no squares. Since $d^{\frac{1}{2}} f_d = h_d \log \epsilon_d$, the statement (1) follows from (18), for $r = 0$; on the other hand, we get (2) by using (18), for $r = 0, 1$, and adding the two formulas.

4. The analogue of (10) for positive primitive quadratic forms with discriminant d is the formula

$$(19) \qquad |\, d \,|^{-\frac{1}{2}} h_d \frac{2\pi}{w_d} = \sum_{n=1}^{\infty} \left(\frac{d}{n}\right) n^{-1},$$

where $w_{-3} = 6$, $w_{-4} = 4$, $w_d = 2$ for $d < -4$. Using exactly the same method as in the preceding section, we get

$$(20) \qquad \sum_{\substack{k \leqq N \\ -k \equiv r \pmod 4}} k^{-\frac{1}{2}} h_{-k} = \frac{\pi N}{24\zeta(3)}\left(1 + \frac{2r-1}{7}\right) + O(N^{\frac{1}{2}} \log N) \qquad (r = 0, 1),$$

$$(21) \qquad \sum_{k \leqq N} h_{-4k} \sim \frac{4\pi}{21\zeta(3)} N^{3/2},$$

$$(22) \qquad \sum_{k \leqq N} h_{-k} = \frac{\pi}{18\zeta(3)} N^{3/2} + O(N \log N),$$

corresponding to (17), (1), (2). The second of these three formulas was already stated by Gauss, without proof.

Let H_d denote the number of classes of all positive quadratic forms $Q(x, y) = ax^2 + bxy + cy^2$ with integral coefficients a, b, c and discriminant $b^2 - 4ac = d$; then $H_d = \sum_t h_t$, where t runs through all divisors of d such that dt^{-1} is a square. By (22),

$$(23) \qquad \sum_{k \leqq N} H_{-k} = \frac{\pi}{18} N^{3/2} + O(N \log N);$$

vice versa, (22) again follows from (23) by an application of the Möbius inversion formula.

There exists in every class of positive quadratic forms $Q(x, y)$ a reduced form,

satisfying the condition $|b| \leqq a \leqq c$; the reduced form is unique whenever $|b| < a < c$. This remark suggests the idea of proving (23) directly, without using the class number formula (19), by a computation of the lattice points in the three-dimensional domain $|\eta| < \xi < \zeta$, $0 < 4\xi\zeta - \eta^2 < T$, for sufficiently large T. In this manner (21) was already proved by Mertens.[5] It is possible to apply a corresponding idea also in the indefinite case; this leads to the quadruple integral J introduced in the next section.

5. Let $Q = Q(x, y) = ax^2 + bxy + cy^2$ be an indefinite quadratic form with arbitrary real coefficients, $a \neq 0$, $b^2 - 4ac = D > 0$. If ρ_1, ρ_2 denote the roots of the quadratic equation $Q(x, 1) = 0$, then $Q = a(x - \rho_1 y)(x - \rho_2 y)$ and $D = a^2(\rho_1 - \rho_2)^2$. We determine the order of ρ_1, ρ_2 by the condition $a(\rho_1 - \rho_2) = \sqrt{D}$, with the positive sign of the square root; then $\rho_1 > \rho_2$ for $a > 0$.

Let λ be a positive parameter and define the positive quadratic form

$$(24)\quad P = P(x, y) = |a|\,\{\lambda^{-1}(x - \rho_1 y)^2 + \lambda(x - \rho_2 y)^2\} = \alpha x^2 + 2\beta xy + \gamma y^2.$$

If $\tau = \xi + i\eta$ denotes the root of $P(x, 1) = 0$ with positive imaginary part, then $\dfrac{\tau - \rho_1}{\tau - \rho_2} = \pm i\lambda$ is pure imaginary; consequently, for variable λ, these points τ describe the half-circle H through ρ_1, ρ_2 in the upper half-plane, with the equation

$$(25)\qquad a(\xi^2 + \eta^2) + b\xi + c = 0.$$

Plainly λ is the tangent of the angle between the diameter from ρ_2 to ρ_1 and the segment from ρ_2 to τ.

The given indefinite quadratic form Q is called reduced, when and only when there exists at least one $\lambda > 0$ such that P is reduced; this means that $2|\beta| \leqq \alpha \leqq \gamma$, or in other words, that τ lies in the fundamental domain F of the modular group, defined by the inequalities $-\frac{1}{2} \leqq \xi \leqq \frac{1}{2}$, $\xi^2 + \eta^2 \geqq 1$. The intersection of H and F is an arc A; consequently, for every given Q, the set of all λ with reduced P is either an interval $\lambda_1 \leqq \lambda \leqq \lambda_2$ or empty. We introduce the non-Euclidean line element $ds = \eta^{-1}(d\xi^2 + d\eta^2)^{\frac{1}{2}}$; then the length of A is

$$(26)\qquad \mu = \mu(a, b, c) = \int_{\lambda_1}^{\lambda_2} \frac{d\lambda}{\lambda} = \log \frac{\lambda_2}{\lambda_1},$$

if A is not empty, and $\mu = 0$ otherwise.

Consider the integral

$$J = \iiint\limits_{D<1} \mu \, da\, db\, dc,$$

extended over the domain of all a, b, c with $a > 0$ and $0 < b^2 - 4ac = D < 1$.

[5] F. Mertens, *Ueber einige asymptotische Gesetze der Zahlentheorie*, Journal für die reine und angewandte Mathematik 77, pp. 289–338 (1874).

In view of the definition (26) of μ we have

$$J = \int_0^\infty \left(\iiint_{\tau \in F} da\, db\, dc \right) \frac{d\lambda}{\lambda};$$

for any given $\lambda > 0$ the inner integral designates the volume of the domain R_λ of all a, b, c with $a > 0$, $0 < D < 1$ and τ lying in F. Instead of a, b, c we introduce by (24) the new variables of integration α, β, γ. If λ is fixed, then P is uniquely determined by Q. On the other hand, it follows from the geometrical meaning of λ and τ, that the roots ρ_1, ρ_2 are uniquely determined by P; since $\alpha\gamma - \beta^2 = D$ and $a > 0$, the indefinite form Q is again uniquely determined by P, when λ is fixed. Consequently the domain R_λ is mapped onto the domain G of all reduced P with $\alpha\gamma - \beta^2 < 1$. This domain G is independent of λ.

By (24), $a^{-1}\alpha = \lambda^{-1} + \lambda$, $a^{-1}\beta = -\lambda^{-1}\rho_1 - \lambda\rho_2$, $a^{-1}\gamma = \lambda^{-1}\rho_1^2 + \lambda\rho_2^2$, whence

$$(27) \qquad d\alpha = (\lambda^{-1} + \lambda)da, \qquad \frac{d(a^{-1}\beta, a^{-1}\gamma)}{d(\rho_1, \rho_2)} = \begin{vmatrix} -\lambda^{-1} & -\lambda \\ 2\lambda^{-1}\rho_1 & 2\lambda\rho_2 \end{vmatrix} = 2(\rho_1 - \rho_2);$$

moreover $a^{-1}b = -\rho_1 - \rho_2$, $a^{-1}c = \rho_1\rho_2$,

$$(28) \qquad \frac{d(a^{-1}b, a^{-1}c)}{d(\rho_1, \rho_2)} = \begin{vmatrix} -1 & -1 \\ \rho_2 & \rho_1 \end{vmatrix} = \rho_2 - \rho_1.$$

From (27) and (28) we get the value of the Jacobian

$$\frac{d(\alpha, \beta, \gamma)}{d(a, b, c)} = -2(\lambda^{-1} + \lambda).$$

Therefore

$$(29) \qquad J = \tfrac{1}{2} \int_0^\infty \frac{d\lambda}{\lambda^2 + 1} \int_G d\alpha\, d\beta\, d\gamma = \frac{\pi}{4} V,$$

where V denotes the volume of G. In order to calculate V, we choose the new variables of integration D, ξ, η, instead of α, β, γ. Since $\alpha^{-1}\beta = -\xi$, $\alpha^{-1}\gamma = \xi^2 + \eta^2$, $D = \alpha\gamma - \beta^2 = (\alpha\eta)^2$, we obtain

$$\frac{d(\alpha^{-1}\beta, \alpha^{-1}\gamma)}{d(\xi, \eta)} = -2\eta, \qquad 2\alpha\eta d(\alpha\eta) = dD,$$

$$\frac{d(\alpha, \beta, \gamma)}{d(D, \xi, \eta)} = -\alpha\eta^{-1} = -D^{\frac{1}{2}}\eta^{-2},$$

$$(30) \qquad V = \int_0^1 D^{\frac{1}{2}}\, dD \int_F \frac{d\xi\, d\eta}{\eta^2} = \frac{2}{3} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{d\xi}{\sqrt{1 - \xi^2}} = \frac{2\pi}{9}.$$

By (29) and (30),

$$(31) \qquad J = \frac{\pi^2}{18}.$$

6. The half-circle H passes through F when and only when one of the two vertices $\frac{1}{2}(\pm 1 + \sqrt{-3})$ of F belongs to the half-circle domain $\xi^2 + \eta^2 + a^{-1}b\xi + a^{-1}c \leqq 0, \eta \geqq 0$. Therefore a reduced form Q is defined by the condition $a + c \leqq \frac{1}{2} | b |$, in case $a > 0$. Since

$$(| b | - a)^2 + 3a^2 = \tfrac{1}{4}(4a - | b |)^2 + \tfrac{3}{4}b^2 = D + 4a(a + c - \tfrac{1}{2} | b |),$$

we obtain the inequalities

$$a^2 \leqq \tfrac{1}{3}D, \qquad b^2 \leqq \tfrac{4}{3}D, \qquad 4a\,| c | = | b^2 - D | \leqq D, \tag{32}$$

for any reduced Q with $a > 0$. Moreover, the positive form P is reduced for all λ in the interval $\lambda_1 \leqq \lambda \leqq \lambda_2$, and then $\frac{3}{4}\alpha^2 = \alpha^2 - \frac{1}{4}\alpha^2 \leqq \alpha\gamma - \beta^2 = D$,

$$a(\lambda^{-1} + \lambda) = \alpha \leqq 2\sqrt{\frac{D}{3}} \qquad (\lambda_1 \leqq \lambda \leqq \lambda_2). \tag{33}$$

For any $\vartheta > 0$ let P_ϑ denote the domain of all reduced Q satisfying the conditions $a \geqq \vartheta$, $D \leqq 1$; in view of (32), P_ϑ is closed and bounded; by (26) and (33), the function μ is continuous in P_ϑ. Define

$$J_\vartheta = \iiint_{P_\vartheta} \mu \, da \, db \, dc,$$

then

$$J = \lim_{\vartheta \to 0} J_\vartheta. \tag{34}$$

On the other hand, $\mu(qa, qb, qc) = \mu(a, b, c)$ for any $q \neq 0$; therefore

$$J_\vartheta = \lim_{N \to \infty} N^{-3} \sum_{Q \in S_1} \mu(a, b, c), \tag{35}$$

where the summation is carried over the set S_1 of all reduced Q with integral coefficients satisfying the conditions $a \geqq \vartheta N$, $b^2 - 4ac = D \leqq N^2$.

We consider first all Q in S_1 whose discriminant $D = h^2$ is a square. Then $4ac = (b + h)(b - h)$ and $0 < h \leqq N$; by (32), $a^2 \leqq \frac{1}{3}N^2$, $b^2 \leqq \frac{4}{3}N^2$. For any pair b, h with $b \neq \pm h$, the number of divisors of $(b + h)(b - h)$ is $o(N)$; on the other hand, in case $b = \pm h$, we have $c = 0$ and $a^2 \leqq \frac{1}{3}N^2$. Consequently the number of all these Q is $N^2 o(N) + NO(N) = o(N^3)$. Since $\mu(a, b, c)$ is bounded, for any fixed ϑ, uniformly with respect to N, it follows that (35) holds good if we cancel in S_1 all Q with $D = h^2$, $h = 1, 2, \cdots, N$.

Next consider the set S_0 of all reduced Q with integral a, b, c satisfying $0 < a < \vartheta N$, $0 < D \leqq N^2$. The points of the arc A attain the minimum of their imaginary part η at one of the end points of A; let $\xi_0 + i\eta_0$ be this end point. On the half-circle H we have

$$\frac{d\lambda}{\lambda} = \frac{d\tau}{\tau - \rho_1} - \frac{d\tau}{\tau - \rho_2} = \frac{(\rho_1 - \rho_2)d\tau}{(\tau - \rho_1)(\tau - \rho_2)}$$

$$= \frac{(\rho_1 - \rho_2)(2\tau - \rho_1 - \rho_2)d\xi}{(\tau - \rho_1)(\tau - \rho_2)(\tau - \bar{\tau})} = \frac{2(\rho_1 - \rho_2)d\xi}{(\tau - \bar{\tau})^2} = \frac{-D^{\frac{1}{2}}\,d\xi}{2a\eta^2};$$

hence

$$\mu(a, b, c) \leqq \frac{D^{\frac{1}{2}}}{2a\eta_0^2}\int_{-\frac{1}{2}}^{\frac{1}{2}} d\xi \leqq \frac{N}{2a\eta_0^2}. \tag{36}$$

On the other hand, by (33),

$$|\log \lambda| < \log \frac{2}{a}\sqrt{\frac{D}{3}} < \log \frac{2N}{a} \qquad (\lambda_1 \leqq \lambda \leqq \lambda_2)$$

$$\mu(a, b, c) < 2 \log \frac{2N}{a}. \tag{37}$$

For any integral k the number of integers a in the interval $(\vartheta N)/2^{k+1} \leqq a < (\vartheta N)/2^k$ is less than $(\vartheta N)/2^k$. Since $b^2 \leqq (4/3)N^2$, we obtain, by (37),

$$\sum_{\substack{Q \in S_0 \\ |c| < 2N}} \mu(a, b, c) = O(N^2) \sum_{k=0}^{\infty} \frac{\vartheta N}{2^k} \log (2^{k+2}\vartheta^{-1}) = \vartheta \log \vartheta^{-1} O(N^3), \tag{38}$$

uniformly in ϑ. Moreover, by (25) and (32),

$$a\eta_0^2 = -a\xi_0^2 - b\xi_0 - c \leqq \tfrac{1}{2}|b| - c < N - c,$$

whence $c < N$, and

$$a\eta_0^2 \geqq -\tfrac{1}{4}a - \tfrac{1}{2}|b| - c > -N - c \geqq \tfrac{1}{2}|c|,$$

provided $|c| \geqq 2N$. In virtue of (32) and (36),

$$\sum_{\substack{Q \in S_0 \\ |c| \geqq 2N}} \mu(a, b, c) = O(N^2) \sum_{a,c} c^{-1}, \tag{39}$$

where a, c run through all integral solutions of $0 < a < \vartheta N$, $4ac \leqq N^2$, $c \geqq 2N$. Obviously,

$$\sum_{a,c} c^{-1} = \sum_{a < \vartheta N} O\left(\log \frac{N}{a}\right) = \vartheta \log \vartheta^{-1} O(N). \tag{40}$$

By (38), (39), (40),

$$N^{-3} \sum_{Q \in S_0} \mu(a, b, c) = \vartheta \log \vartheta^{-1} O(1). \tag{41}$$

Finally, let S be the set of all reduced Q with integral a, b, c such that the discriminant $b^2 - 4ac \leqq N^2$ and not a square number. If Q belongs to S, then also $-Q$; moreover, $a \neq 0$. By (35) and (41),

$$\limsup_{N\to\infty} |2J_\vartheta - N^{-3} \sum_{Q \in S} \mu(a, b, c)| = \vartheta \log \vartheta^{-1} O(1). \tag{42}$$

If $\vartheta \to 0$, then $\vartheta \log \vartheta^{-1} \to 0$ and $J_\vartheta \to J$, by (34). Since S is independent of ϑ, we infer from (31) and (42) that

$$\lim_{N\to\infty} N^{-3} \sum_{Q \in S} \mu(a, b, c) = 2J = \frac{\pi^2}{9}. \tag{43}$$

7. Let Q be primitive, i.e., $(a, b, c) = 1$. It is well known that the matrix $\mathfrak{M}$ of ever linear transformation $x \to px + qy$, $y \to rx + sy$ of Q into itself, with integral p, q, r, s and $ps - qr = +1$, has the form

$$\mathfrak{M} = \pm\begin{pmatrix} \frac{t + bu}{2} & au \\ -cu & \frac{t - bu}{2} \end{pmatrix}^{l} \qquad (l = 0, \pm 1, \pm 2, \cdots),$$

where t, u is the smallest positive integral solution of $t^2 - Du^2 = 4$. The corresponding modular substitution has the fixed points ρ_1, ρ_2 and leaves the half-circle H invariant. A fundamental domain on H for the cyclic group of these modular substitutions is given by any arc B on H with the non-Euclidean length $2 \log \epsilon_D$, where $\epsilon_D = \dfrac{t + u\sqrt{D}}{2}$.

Let $Q_k = a_k x^2 + b_k xy + c_k y^2$ $(k = 1, \cdots, g)$ denote all reduced forms which are equivalent, in the narrow sense, with the given primitive form Q, and let A_k denote the arc A for the particular form Q_k. It follows from the definition of A and B, that the arcs $A_1, \cdots, A_g$ are equivalent to certain arcs on the half-circle H which cover B without gaps and overlappings. Consequently,

$$\sum_{k=1}^{g} \mu(a_k, b_k, c_k) = 2 \log \epsilon_D .$$

Summing over all primitive Q with given discriminant D, we obtain

$$\sum_{\substack{b^2 - 4ac = D \\ (a,b,c)=1}} \mu(a, b, c) = 2h_D \log \epsilon_D , \tag{44}$$

where h_D denotes the class number.

By (43) and (44),

$$\sum_{Dq^2 \leq N^2} h_D \log \epsilon_D \sim \frac{\pi^2}{18} N^3, \tag{45}$$

where D, q run through all positive integral solutions of $Dq^2 \leq N^2$ and D is not a square. Denoting the left-hand member in (45) by $g(N)$, we get

$$\sum_{D \leq N^2} h_D \log \epsilon_D = \sum_{t \leq N} \mu(t) g(t^{-1}N) \sim \frac{\pi^2}{18} \sum_{t \leq N} \mu(t)\left(\frac{N}{t}\right)^3 \sim \frac{\pi^2}{18\zeta(3)} N^3.$$

This result differs from (2) only in the estimation of the error term, viz., $o(N^3)$ instead of $O(N^2 \log N)$. It is possible to obtain this better appraisal of the error term by defining the parameter ϑ of §6 in a suitable way as a function of N.

In order to prove also formula (1) by our second method, we cancel in the sum on the right-hand side of (35) all terms in which either b is odd or a, b, c are all even. The remaining triplets a, b, c constitute exactly 3 systems of residue classes modulo 2; consequently we have then to multiply the left-hand member

in (35) by the factor 3/8. Since $b^2 - 4ac = D = 4k$ is divisible by 4 and (a, b, c) is odd, we obtain instead of (45) the relationship

$$\sum_{4kq^2 \leq N^2} h_{4k} \log \epsilon_{4k} \sim \frac{\pi^2}{48} N^3,$$

where k, q run through all positive integral solutions of $4kq^2 \leq N^2$ with odd q and k is not a square. It follows that

$$\sum_{k \leq N^2} h_{4k} \log \epsilon_{4k} \sim \frac{\pi^2}{48} \sum_{q \leq N} \mu(q) \left(\frac{2N}{q}\right)^3 \sim \frac{4\pi^2}{21\zeta(3)} N^3,$$

and this is the formula of Gauss.

II. Forms of m Variables

8. Let $\mathfrak{S}$ be a real symmetric matrix of signature n, $m - n$, and let the absolute value of the determinant $|\mathfrak{S}|$ be $S > 0$. We denote by H the set of all positive real symmetric $\mathfrak{H}$ which satisfy the condition $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$. It is known that H is a manifold of $n(m - n)$ dimensions with the parametric representation

$$\mathfrak{H} = 2\mathfrak{W}^{-1}[\mathfrak{X}'\mathfrak{S}] - \mathfrak{S}, \quad \mathfrak{W} = \mathfrak{S}[\mathfrak{X}] > 0, \tag{46}$$

where $\mathfrak{X}$ is a variable real matrix with m rows, n columns, rank n and positive $\mathfrak{S}[\mathfrak{X}]$; two such matrices $\mathfrak{X} = \mathfrak{X}_1$, $\mathfrak{X}_2$ represent the same point $\mathfrak{H}$ of H, when and only when $\mathfrak{X}_2 = \mathfrak{X}_1\mathfrak{R}$ with real n-rowed $\mathfrak{R}$.

Let F be the domain of the reduced positive quadratic forms of m variables, in the definition of Minkowski. Moreover, let W be a bounded domain in the space of the positive quadratic forms of n variables and denote by X the set of all $\mathfrak{X}$ which fulfill the two conditions $\mathfrak{W} \in W$, $\mathfrak{H} \in F$. We define

$$\rho_0 = 1, \qquad \rho_k = \prod_{l=1}^{k} \pi^{k/2}/\Gamma\left(\frac{k}{2}\right) \qquad (k = 1, 2, \cdots)$$

and

$$\mu(\mathfrak{S}) = \rho_{m-n} S^{m/2} \int_X |\mathfrak{W}|^{(n-m+1)/2} d\mathfrak{X} \Big/ \int_W d\mathfrak{W}. \tag{47}$$

Let $\mathfrak{R}$ be the matrix formed by the last n rows of $\mathfrak{X}$ and put $\mathfrak{X}\mathfrak{R}^{-1} = \begin{pmatrix}\mathfrak{Y}\\ \mathfrak{E}\end{pmatrix} = \mathfrak{Z}$, $\mathfrak{S}[\mathfrak{Z}] = \mathfrak{U}$, then $\mathfrak{H} = 2\mathfrak{U}^{-1}[\mathfrak{Z}'\mathfrak{S}] - \mathfrak{S}$, $\mathfrak{U} = \mathfrak{S}[\mathfrak{Z}] > 0$, and

$$\int_X |\mathfrak{W}|^{(n-m+1)/2} d\mathfrak{X} \Big/ \int_W d\mathfrak{W} = \rho_n \int_Y |\mathfrak{U}|^{-m/2} d\mathfrak{Y},$$

where Y is the set of all $\mathfrak{Y}$ satisfying $\mathfrak{U} > 0$, $\mathfrak{H} \in F$. This proves that $\mu(\mathfrak{S})$ is independent of W. In the particular case of an indefinite binary form $\mathfrak{S}[\mathfrak{x}] = ax_1^2 + bx_1x_2 + cx_2^2$ it is easily seen that $2\mu(\mathfrak{S})$ has the value μ in (26), provided F is defined with respect to the narrow unimodular group; however, we shall now assume that equivalence is understood in the wide sense.

We set

$$J = \int_{S<1} \mu(\mathfrak{S}) S^{(1-m)/2}\, d\mathfrak{S}, \tag{48}$$

the integration extended over the space of all $\mathfrak{S}$ with $0 < S < 1$ and given signature n, $m - n$. By (47),

$$J \int_W d\mathfrak{W} = \rho_{m-n} \int_R \left\{ \int (S \mid \mathfrak{W} \mid)^{(n-m+1)/2}\, d\mathfrak{S} \right\} d\mathfrak{X}; \tag{49}$$

the outer integration is carried over the space R of all real $\mathfrak{X}$ with rank n, the inner integration is restricted by the three conditions $0 < S < 1$, $\mathfrak{W} \in W$, $\mathfrak{H} \in F$, for any given $\mathfrak{X}$ in R.

By the substitution (46) we introduce the new variable of integration $\mathfrak{H}$ instead of $\mathfrak{S}$, for fixed $\mathfrak{X}$. Since $\mathfrak{S}\mathfrak{X} = \mathfrak{H}\mathfrak{X}$, we obtain

$$\mathfrak{H}[\mathfrak{X}] = \mathfrak{W} > 0, \qquad \mathfrak{S} = 2\mathfrak{W}^{-1}[\mathfrak{X}'\mathfrak{H}] - \mathfrak{H}; \tag{50}$$

vice versa, (46) follows from (50). Consequently (46) defines a birational mapping of the space of all $\mathfrak{S}$ with the given signature onto the space of the positive $\mathfrak{H}$. In view of the relationship $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$, we can determine a real matrix $\mathfrak{C}$ such that $\mathfrak{H}[\mathfrak{C}] = \mathfrak{E}$ and $\mathfrak{S}[\mathfrak{C}] = \begin{pmatrix} \mathfrak{E}^{(n)} & 0 \\ 0 & -\mathfrak{E}^{(m-n)} \end{pmatrix} = \mathfrak{S}_0$, say, for any given pair $\mathfrak{S}$, $\mathfrak{H}$. This proves that the Jacobian $d\mathfrak{S}/d\mathfrak{H} = j$ depends only upon m and n; therefore it suffices to compute j at the particular point $\mathfrak{H} = \mathfrak{E}$, $\mathfrak{S} = \mathfrak{S}_0$. It follows from the formula $\mathfrak{S}\mathfrak{X} = \mathfrak{H}\mathfrak{X}$ that then all elements of the last n rows of the fixed matrix $\mathfrak{X}$ are 0. Let $\mathfrak{S} = \begin{pmatrix} \mathfrak{S}_1^{(n)} & \mathfrak{S}_{12} \\ \mathfrak{S}_{21} & \mathfrak{S}_2 \end{pmatrix}$, then

$$\mathfrak{H} = \begin{pmatrix} \mathfrak{S}_1 & \mathfrak{S}_{12} \\ \mathfrak{S}_{21} & 2\mathfrak{S}_1^{-1}[\mathfrak{S}_{12}] - \mathfrak{S}_2 \end{pmatrix}, \text{ by (46), and } d\mathfrak{H} = \begin{pmatrix} d\mathfrak{S}_1 & d\mathfrak{S}_{12} \\ d\mathfrak{S}_{21} & -d\mathfrak{S}_2 \end{pmatrix}$$

for $\mathfrak{S}_{12} = 0$, whence $j = (-1)^{((m-n)(m-n+1))/2}$.

By (49), (50),

$$J \int_W d\mathfrak{W} = \rho_{m-n} \int_{\substack{F \\ |\mathfrak{H}|<1}} |\mathfrak{H}|^{(n-m+1)/2} \left(\int_{\mathfrak{H}[\mathfrak{X}] \in W} |\mathfrak{W}|^{(n-m+1)/2}\, d\mathfrak{X} \right) d\mathfrak{H};$$

on the other hand,

$$\int_{\mathfrak{H}[\mathfrak{X}] \in W} |\mathfrak{W}|^{(n-m+1)/2}\, d\mathfrak{X} = \frac{\rho_m}{\rho_{m-n}} |\mathfrak{H}|^{-(n/2)} \int_W d\mathfrak{W}.$$

Consequently, using Minkowski's formula for the volume of the domain of the reduced $\mathfrak{H}$ with $|\mathfrak{H}| < 1$, we obtain

$$J = \rho_m \int_{\substack{F \\ |\mathfrak{H}|<1}} |\mathfrak{H}|^{(1-m)/2}\, d\mathfrak{H} = \prod_{k=2}^{m} \zeta(k). \tag{51}$$

9. Let $s_{11} = s$ be the first diagonal element of $\mathfrak{S} = (s_{kl})$. For any $\vartheta > 0$ we define

$$J_\vartheta = \int\limits_{\substack{\text{abs } s > \vartheta \\ 0 < S < 1}} \mu(\mathfrak{S}) S^{(1-m)/2} \, d\mathfrak{S}.$$

By (48),

$$J = \lim_{\vartheta \to 0} J_\vartheta. \tag{52}$$

The real symmetric matrix $\mathfrak{S}$ is called reduced whenever there exists a solution $\mathfrak{H}$ of $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$ in the domain F. If $\mu(\mathfrak{S}) \neq 0$, then necessarily $\mathfrak{S}$ is reduced, in view of definition (47). Let $h_1, \cdots, h_m$ be the diagonal elements of $\mathfrak{H}$; since $\mathfrak{H} + \mathfrak{S}$ and $\mathfrak{H} - \mathfrak{S}$ are non-negative, it follows that $s_{kl}^2 \leqq h_k h_l$ $(k, l = 1, \cdots, m)$. If $\mathfrak{H}$ lies in F, then $h_1 \cdots h_m S^{-m}$ and the $m - 1$ ratios $h_1/h_2, \cdots, h_{m-1}/h_m$ are bounded. Consequently, if moreover $s^2 > \vartheta^2$ and $0 < S < 1$, then $\mathfrak{H}$, $\mathfrak{S}$ and S^{-1} are bounded, for any given ϑ, and $\mu(\mathfrak{S})$ is continuous, by (47). Since $\mu(q\mathfrak{S}) = \mu(\mathfrak{S})$, for all scalar factors $q > 0$, we infer from the definition of an integral that

$$J_\vartheta = \lim_{N \to \infty} N^{-1} \sum_{\substack{\text{abs } s > \vartheta N^{1/m} \\ S \leqq N}} \mu(\mathfrak{S}) S^{(1-m)/2}, \tag{53}$$

where $\mathfrak{S}$ runs through all integral symmetric matrices of signature n, $m - n$ with abs $s > \vartheta N^{1/m}$ and $S \leqq N$. Performing the passage to the limit $\vartheta \to 0$, we obtain, by (52) and (53),

$$J \leqq \liminf_{N \to \infty} N^{-1} \sum_{S \leqq N} \mu(\mathfrak{S}) S^{(1-m)/2}. \tag{54}$$

For any given integral $\mathfrak{S}$ there exist only a finite number of equivalent reduced matrices $\mathfrak{S}_1, \cdots, \mathfrak{S}_g$. It is known that

$$2\rho(\mathfrak{S}) S^{(m+1)/2} = \sum_{k=1}^{g} \mu(\mathfrak{S}_k), \tag{55}$$

where $\rho(\mathfrak{S})$ is defined by (3), except in the case $m = 2$, $n = 1$, S a square. By (51), (54), (55),

$$\liminf_{N \to \infty} N^{-1} \sum_{S \leqq N} S\rho(\mathfrak{S}) \geqq \frac{1}{2} \prod_{k=2}^{m} \zeta(k); \tag{56}$$

in the left-hand member $\mathfrak{S}$ runs over a system of representatives of all classes with signature n, $m - n$ whose determinants have the absolute values $S \leqq N$.

In our proof of (56) we have tacitly assumed, in some places, that $0 < n < m$. However, in the definite case $n = 0$ or m, formula (56) follows directly from (5), (51) and the definition of an integral.

If it is true that the two passages to the limit $N \to \infty$, $\vartheta \to 0$ can be interchanged in the right-hand member of (53), then we should obtain from (53)

statement (4) of the Theorem, instead of the much weaker result (56). However, the direct proof of the legitimacy of this interchange seems to be very complicated in the general indefinite case. For binary forms, we overcame this difficulty in §6 by using the estimates (36) and (37) for μ; but the corresponding way becomes impractical in case $m \geqq 3$. In order to get the desired result we shall also apply the generalization of the class number formulas (10) and (19), viz., the formula for the measure of a genus. Therefore our proof of the Theorem will be a combination of the ideas of the two different proofs given in the first Chapter.

The Theorem is trivial in case $m = 1$; for $m = 2$, it is easily derived from (17) and (20), even with the estimate $O(N^{\frac{1}{2}} \log N)$ of the error term. For the rest of the paper we shall assume that $m \geqq 3$.

10. Let $\mathfrak{S}_1, \cdots, \mathfrak{S}_h$ denote a complete system of representatives of the classes in the genus of $\mathfrak{S}$ and define

$$\nu(\mathfrak{S}) = \sum_{k=1}^{h} \rho(\mathfrak{S}_k). \tag{57}$$

It is known that

$$\nu(\mathfrak{S}) = 2 \prod_p \frac{2q^{m(m-1)/2}}{E_q(\mathfrak{S})}, \tag{58}$$

where p runs through all primes, $q = p^t$ is a sufficiently large power of p and $E_q(\mathfrak{S})$ is the order of the group of units of $\mathfrak{S}$ modulo q. If p^b is the highest power of p dividing $2S$, then t may be any integer $> 2b$; in particular, for $b = 0$, i.e., $(p, 2S) = 1$, we have the explicit formula

$$\tfrac{1}{2} q^{-m(m-1)/2} E_q(\mathfrak{S}) = \begin{cases} \prod_{k=1}^{(m-1)/2} (1 - p^{-2k}) & (m \text{ odd}) \\ (1 - \delta p^{-m/2}) \prod_{k=1}^{m/2-1} (1 - p^{-2k}) & (m \text{ even}), \end{cases} \tag{59}$$

where δ denotes the Legendre symbol $\left(\frac{(-1)^{m/2} |\mathfrak{S}|}{p}\right)$.

Let $N \geqq 8$, $S \leqq N$ and

$$Q = \prod_p p^{3[\log N/\log p]}; \tag{60}$$

then $p^b \leqq 2S \leqq 2N$, whence $b \leqq \left[\frac{\log N}{\log p}\right] + \left[\frac{\log 2}{\log p}\right] < \frac{3}{2}\left[\frac{\log N}{\log p}\right]$ and $S \mid Q$. Defining

$$\alpha_0(\mathfrak{S}) = \prod_{p>N} \frac{2p^{m(m-1)/2}}{E_p(\mathfrak{S})}, \qquad \alpha(\mathfrak{S}) = \frac{2^{\pi(N)} Q^{m(m-1)/2}}{E_Q(\mathfrak{S})}, \tag{61}$$

where $\pi(N)$ denotes the number of primes $\leqq N$, we obtain

(62) $$\nu(\mathfrak{S}) = 2\alpha_0(\mathfrak{S})\alpha(\mathfrak{S}), \qquad \alpha_0(\mathfrak{S}) = 1 + O(N^{-1}),$$

by (58), (59) and (60).

On the other hand, the number of modulo Q incongruent integral m-rowed matrices $\mathfrak{G}$ with $(|\mathfrak{G}|, Q) = 1$ is

(63) $$A = A_Q = Q^{m^2} \prod_{p \leq N} \prod_{k=1}^{m} (1 - p^{-k}).$$

Obviously the ratio $A/E_Q(\mathfrak{S})$ is the number of symmetric $\mathfrak{S}_0$ modulo Q which are equivalent to $\mathfrak{S}$ modulo Q. For these $\mathfrak{S}_0$ we have

$$|\mathfrak{S}_0| \equiv |\mathfrak{S}| x^2 (\mathrm{mod}\ Q), \qquad (x, Q) = 1;$$

it follows that the number of determinants $|\mathfrak{S}_0|$ modulo Q equals the number of quadratic residues x^2 modulo $S^{-1}Q$ which are prime to Q; this number has the value

(64) $$B = 2^{-1-\tau(N)} S^{-1} Q \prod_{p \leq N} (1 - p^{-1}).$$

Consequently the ratio $A/BE_Q(\mathfrak{S})$ is the number of symmetric $\mathfrak{S}_0$ modulo Q fulfilling the two conditions $\mathfrak{S}_0 \sim \mathfrak{S}$ (mod Q), $|\mathfrak{S}_0| \equiv |\mathfrak{S}|$ (mod Q).

Finally, let $\mathfrak{S}$ run through a system of representatives of all genera with signature $n, m - n$ and given determinant $|\mathfrak{S}| = (-1)^{m-n}S$. We denote by $C_n(S)$ the number of residue classes modulo Q which contain an integral symmetric matrix with this signature and determinant. By (61), (63), (64),

(65) $$2Q^{m(m+1)/2-1} S \prod_{p \leq N} \prod_{k=2}^{m} (1 - p^{-k}) \sum_{\mathfrak{S}} \alpha(\mathfrak{S}) = \frac{A}{B} \sum_{\mathfrak{S}} \frac{1}{E_Q(\mathfrak{S})} = C_n(S).$$

11. By the reciprocity formula for Gaussian sums we have

$$\sum_{\mathfrak{r}\,(\mathrm{mod}\ Q)} e^{2\pi i Q^{-1}\mathfrak{S}[\mathfrak{r}]} = \epsilon^{2n-m} S^{\frac{1}{2}} (2Q)^{m/2}, \qquad \epsilon = e^{\pi i/4}.$$

Define $n_0 = n - 2$ for $n \geq 2$ and $n_0 = n + 2$ for $n < 2$; then $0 \leq n_0 \leq m$, in view of $m \geq 3$ and $0 \leq n \leq m$. Since $\epsilon^{2n_0} = -\epsilon^{2n}$ and $(-1)^{n_0} = (-1)^n$, it follows that two symmetric matrices with the same determinant $(-1)^{m-n}S$ and the different signatures $n, m - n$ and $n_0, m - n_0$ never belong to the same residue class modulo Q. Consequently the sum $C_n(S) + C_{n_0}(S)$ is at most equal to the number $D(S)$ of modulo Q incongruent integral symmetric matrices $\mathfrak{S}_1$ satisfying $|\mathfrak{S}_1| \equiv (-1)^{m-n}S(\mathrm{mod}\ Q)$.

Let $\mathfrak{S}$ and $\mathfrak{S}_0$ run through a system of representatives of all classes of integral symmetric matrices of signatures $n, m - n$ and $n_0, m - n_0$ whose determinants have an absolute value $S \leq N$. Since

$$\prod_{p \leq N} \prod_{k=2}^{m} (1 - p^{-k})^{-1} = (1 + O(N^{-1})) \prod_{k=2}^{m} \zeta(k),$$

we infer from (57), (62), (65) the inequality

$$\sum_{S \leq N} \{\rho(\mathfrak{S}) + \rho(\mathfrak{S}_0)\}S < (1 + o(1))Q^{1-m(m+1)/2} \sum_{S=1}^{N} D(S) \prod_{k=2}^{m} \zeta(k). \tag{66}$$

In the remaining sections we shall prove that

$$Q^{1-\frac{1}{2}m(m+1)} \sum_{S=1}^{N} D(S) \sim N \qquad (N \to \infty). \tag{67}$$

Plainly the Theorem is an immediate consequence of (56), (66) and (67).

12. Define

$$e_q(x) = e^{2\pi i(x/q)}, \qquad \beta\left(\frac{l}{q}\right) = q^{-m(m+1)/2} \sum_{\mathfrak{S}(\mathrm{mod}\ q)} e_q(l \mid \mathfrak{S} \mid) \quad (q > 0, (l, q) = 1);$$

then

$$QD(S) = \sum_{l=1}^{Q} e_Q((-1)^{m-n-1} lS) \sum_{\mathfrak{S}(\mathrm{mod}\ Q)} e_Q(l \mid \mathfrak{S} \mid),$$

$$Q^{1-\frac{1}{2}m(m+1)} D(S) = \sum_{q \mid Q} \sum_{\substack{l\,(\mathrm{mod}\ q) \\ (l,q)=1}} \beta\left(\frac{l}{q}\right) e_q((-1)^{m-n-1} lS).$$

Moreover

$$\mathrm{abs} \sum_{S=1}^{N} e_q((-1)^{m-n-1} lS) \leq \mathrm{Min}\left(N, \frac{1}{\left|\sin \frac{\pi l}{q}\right|}\right)$$

and

$$\sum_{l=1}^{q-1} \mathrm{Min}\left(N, \frac{1}{\left|\sin \frac{\pi l}{q}\right|}\right) = qO(\log N).$$

Consequently, in order to prove (67), it suffices to deduce the estimate

$$\sum_{q \mid Q} q\beta_q = o\left(\frac{N}{\log N}\right), \tag{68}$$

where β_q denotes the maximum of the $\varphi(q)$ numbers $\mathrm{abs}\, \beta\left(\frac{l}{q}\right)$ $(l = 1, \cdots, q;\ (l, q) = 1)$.

It is easily seen that

$$\beta_{q_1 q_2} = \beta_{q_1} \beta_{q_2} \qquad ((q_1, q_2) = 1). \tag{69}$$

It remains to determine an upper appraisal of β_q in case $q = p^t (t = 1, 2, \cdots)$.

Let $p \nmid 2a$ and consider all residue classes of symmetric $\mathfrak{S}$ modulo p satisfying

$|\mathfrak{S}| \equiv a (\text{mod } p)$. Any two such $\mathfrak{S}$ are equivalent modulo p; in virtue of (59) and (63), their number has the value

$$W_a = \frac{2A_p}{(p-1)E_p(\mathfrak{S})} = p^{m(m+1)/2-1}(1 + O(p^{-2})).$$

Hence

$$\beta\left(\frac{l}{p}\right) = p^{-m(m+1)/2}\left(p^{m(m+1)/2} + \sum_{a=1}^{p-1} W_a(e_p(la) - 1)\right) = O(p^{-2}),$$

$$p\beta_p = O(p^{-1}), \tag{70}$$

for the set of all primes p.

Next let $q = p^t$, $t \geqq 2$, $b = [(t+1)/2]$, $q_0 = p^b$; let $\mathfrak{S}_1$ and $\mathfrak{S}_2$ run through complete systems of residues modulo q_0 and q/q_0. If $\mathfrak{S}_0 = (s_{kl})$ denotes the adjoint matrix of $\mathfrak{S}_1$, then

$$|\mathfrak{S}_1 + q_0\mathfrak{S}_2| \equiv |\mathfrak{S}_1| + q_0\sigma(\mathfrak{S}_0\mathfrak{S}_2) \qquad (\text{mod } q),$$

whence

$$\beta\left(\frac{l}{q}\right) = q^{-m(m+1)/2} \sum_{\mathfrak{S}_1} e_q(l\,|\,\mathfrak{S}_1\,|) \sum_{\mathfrak{S}_2} e_q(lq_0\,\sigma(\mathfrak{S}_0\mathfrak{S}_2)). \tag{71}$$

The inner sum vanishes, except when q/q_0 is a common divisor of all numbers $s_{kk}, 2s_{kl} (k, l = 1, \cdots, m)$, and then it has the value $(q/q_0)^{m(m+1)/2}$. If the latter case occurs exactly F_q times, then

$$\beta_q \leqq q_0^{-m(m+1)/2} F_q. \tag{72}$$

13. Let p^{α_1} be the highest power of the prime p dividing all elements of $\mathfrak{S}_1$ and q_0; if $\alpha_1 = 0$, then $\mathfrak{S}_1$ is called primitive modulo q_0. Put $q_0 p^{-\alpha_1} = q_1$; plainly, $\mathfrak{S}_1 = p^{\alpha_1}\mathfrak{T}$, where $\mathfrak{T}$ is modulo q_1 primitive and uniquely determined. Suppose $\alpha_1 < b$, i.e., $p \mid q_1$. If all diagonal elements of $\mathfrak{T} = (t_{kl})$ are divisible by p, then there exists an element t_{kl}, with $k \neq l$, which is not a multiple of p, and the same holds for the principal minor $t_{kk}t_{ll} - t_{kl}^2$. Consequently there exist principal minors in $\mathfrak{T}$ which are not divisible by p; let r_1 be the maximal order of these minors. After a suitable permutation in the rows and the columns of $\mathfrak{T}$ we obtain

$$\mathfrak{T}[\mathfrak{C}_1] = \begin{pmatrix} \mathfrak{T}_1 & \mathfrak{T}_{12} \\ \mathfrak{T}_{21} & \mathfrak{T}_2 \end{pmatrix}, \tag{73}$$

where $\mathfrak{C}_1$ is the matrix of the permutation, $\mathfrak{T}_1 = \mathfrak{T}_1^{(r_1)}$, $p \nmid |\mathfrak{T}_1|$. Let $\mathfrak{A}$ be an integral matrix satisfying $\mathfrak{T}_1\mathfrak{A} \equiv \mathfrak{E} \pmod{q_1}$ and set

$$\mathfrak{C}_2 = \begin{pmatrix} \mathfrak{E} & -\mathfrak{A}\mathfrak{T}_{12} \\ 0 & \mathfrak{E} \end{pmatrix}, \tag{74}$$

then

$$\mathfrak{S}_1[\mathfrak{C}_1\mathfrak{C}_2] \equiv p^{\alpha_1}\begin{pmatrix}\mathfrak{T}_1 & 0\\ 0 & \mathfrak{T}_0\end{pmatrix} \pmod{q_0} \tag{75}$$

with integral $\mathfrak{T}_0$.

We denote by p^{ω_l} $(l = 1, \cdots, m)$ the elementary divisors of $\mathfrak{S}_1$ modulo q_0, such that $0 \leqq \omega_1 \leqq \cdots \leqq \omega_m \leqq b$; moreover, let $\alpha_1, \cdots, \alpha_k$ be the different values of the ω_l, where $0 \leqq \alpha_1 < \alpha_2 < \cdots < \alpha_k \leqq b$, and denote by r_l the multiplicity of α_l in the sequence $\omega_1, \cdots, \omega_m$; plainly, α_1 and r_1 have their former meaning. The $(m - r_1)$-rowed matrix $\mathfrak{T}_0$ has modulo q_1 the elementary divisors $p^{\omega_l-\alpha_1}(l \equiv r_1 + 1, \cdots, m)$.

Let $F(q_0; \omega_1, \cdots, \omega_m)$ be the number of modulo q_0 incongruent $\mathfrak{S}_1$ with given $\omega_1, \cdots, \omega_m$, and put $\omega_0 = 0, \gamma_l = \omega_l - \omega_{l-1}$ $(l = 1, \cdots, m)$. We shall prove by induction that

$$q_0^{-m(m+1)/2}F(q_0; \omega_1, \cdots, \omega_m) < c_m p^{-\gamma}, \qquad \gamma = \sum_{l=1}^{m} \frac{l(l+1)}{2}\gamma_{m-l+1}, \tag{76}$$

where c_m depends only upon m. The assertion (76) is empty in case $m = 0$. Using (76) for $m - r_1$, q_1 instead of m, q_0, we obtain

$$q_1^{-(m-r_1)(m-r_1+1)/2}F(q_1; \omega_{r_1+1} - \alpha_1, \cdots, \omega_m - \alpha_1) < c_{m-r_1}p^{-\delta}, \tag{77}$$

where

$$\delta = \sum_{l=1}^{m-r_1} \frac{l(l+1)}{2}\gamma_{m-l+1} = \gamma - \frac{m(m+1)}{2}\alpha_1; \tag{78}$$

moreover, the number of modulo q_1 incongruent $(\mathfrak{T}_1\mathfrak{T}_{12})$ in (73) is at most $q_1^{r_1(m-(r_1-1)/2)}$, the number of permutation matrices $\mathfrak{C}_1$ is $m!$. By (73), (74), (75), (77), (78),

$$q_0^{-m(m+1)/2}F(q_0; \omega_1, \cdots, \omega_m) < m!\, c_{m-r_1}p^{-\delta}(q_1/q_0)^{m(m+1)/2} \leqq c_m p^{-\gamma},$$

with $c_m = m! \operatorname{Max} c_{m-r_1}$; and this is the assertion (76).

Suppose now that $\mathfrak{S}_0 = (s_{kl})$ is the adjoint matrix of $\mathfrak{S}_1$ and that $q/q_0 = p^{t-b}$ is a common factor of all numbers s_{kk}, $2s_{kl}$ $(k, l = 1, \cdots, m)$; then, *a fortiori*, $2p^{b-t}\mathfrak{S}_0$ is integral; hence $t - b \leqq \omega_1 + \cdots + \omega_{m-1}$ for $p \neq 2$ and $t - b \leqq \omega_1 + \cdots + \omega_{m-1} + 1$ for $p = 2$. On the other hand,

$$\omega_1 + \cdots + \omega_{m-1} = \sum_{l=1}^{m}(l-1)\gamma_{m-l+1}$$

and $l(l + 1) - 6(l - 1) = (l - 2)(l - 3) \geqq 0$ $(l = 1, \cdots, m)$, where the equality sign holds only for $l = 2, 3$; therefore $\omega_1 + \cdots + \omega_{m-1} \leqq \frac{1}{3}\gamma$, with the equality sign only when $\gamma_l = 0$ for $l \neq m - 1, m - 2$. It follows that

$$p^{-\gamma} \leqq 8\left(\frac{q_0}{q}\right)^3 = 8q^{-1}p^{3[(t+1)/2]-2t}; \tag{79}$$

in case $t = 3$, $p \neq 2$, we have

$$p^{-\gamma} \leqq \frac{1}{pq}, \tag{80}$$

except when $\gamma_l = 0$ for $l \neq m - 1, m - 2$ and $\gamma_{m-1} + 2\gamma_{m-2} = 1$. Plainly this exception is the case $\omega_l = 0 (l = 1, \cdots, m - 2)$, $\omega_{m-1} = \omega_m = 1$.

14. The number of systems of integers $\omega_1, \cdots, \omega_m$ satisfying $0 \leqq \omega_1 \leqq \cdots \leqq \omega_m \leqq b$ is $< (b + 1)^m$. By (71), (72), (76), (79) and (80), we obtain the inequalities

$$q\beta_q < t^m p^{3[(t+1)/2]-2t} O(1) \qquad (q = p^t;\ t = 2, 3, \cdots), \tag{81}$$

$$q\beta_q < O(p^{-1}) + p^{3-m(m+1)} \operatorname{abs} \sum_{\mathfrak{S}_1} e_q(|\mathfrak{S}_1|) \qquad (q = p^3), \tag{82}$$

where $\mathfrak{S}_1$ runs through all symmetric matrices modulo p^2 with $\omega_1 = \cdots = \omega_{m-2} = 0, \omega_{m-1} > 0$. In order to estimate this last sum, we put $\mathfrak{S}_1 = \mathfrak{R}_1 + p\mathfrak{R}_2$, where $\mathfrak{R}_2$ runs over all residue classes modulo p and $\mathfrak{R}_1$ over all residue classes modulo p with $\omega_{m-2} = 0, \omega_{m-1} = 1$. The number of these $\mathfrak{R}_1$ is $O(p^{m(m+1)/2-3})$, by (76). Let $\mathfrak{R}_1$ be given and determine a unimodular $\mathfrak{C}$, according to (75), such that $\mathfrak{R}_1[\mathfrak{C}] \equiv \begin{pmatrix} \mathfrak{T}_1^{(m-2)} & 0 \\ 0 & p\mathfrak{T}_2 \end{pmatrix} \pmod{p^2}$, $\mathfrak{T}_2 = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ with integral a, b. Put $\mathfrak{R}_2[\mathfrak{C}] = \begin{pmatrix} \mathfrak{W}_1 & \mathfrak{W}_{12} \\ \mathfrak{W}_{21} & \mathfrak{W}_2 \end{pmatrix}$, $\mathfrak{W}_2 = \begin{pmatrix} x & y \\ y & z \end{pmatrix}$, then $|\mathfrak{S}_1| \equiv |\mathfrak{R}_1| + p^2 |\mathfrak{T}_1| \, |\mathfrak{W}_2 + \mathfrak{T}_2| \pmod{p^3}$, $p \nmid |\mathfrak{T}_1|$, $|\mathfrak{W}_2 + \mathfrak{T}_2| = (x + a)(z + b) - y^2$. Obviously,

$$\sum_{x \pmod p} e_q(|\mathfrak{S}_1|) = 0,$$

except when $z \equiv -b \pmod p$. If $\mathfrak{R}_1$ is given, then the number of $\mathfrak{R}_2$ with $z \equiv -b \pmod p$ is $p^{m(m+1)/2-1}$. Consequently,

$$\sum_{\mathfrak{S}_1} e_q(|\mathfrak{S}_1|) = O(p^{m(m+1)-4})$$

and

$$q\beta_q = O(p^{-1}) \qquad (q = p^3), \tag{83}$$

by (82).

It follows from (70), (81), (83) that

$$\sum_{t=1}^{\infty} p^t \beta_{p^t} = O(p^{-1}) \left(1 + \sum_{t=4}^{\infty} t^m p^{3[(t+1)/2]-2t+1}\right) = O(p^{-1}). \tag{84}$$

In view of (60), (69), (84),

$$\sum_{q|Q} q\beta_q < \prod_{p \leqq N} (1 + O(p^{-1})) = e^{O(1) \sum_{p \leqq N} p^{-1}} = e^{O(\log \log N)}$$

$$= (\log N)^{O(1)} = o\left(\frac{N}{\log N}\right),$$

and this is the assertion (68). The proof of the Theorem is now complete.

THE INSTITUTE FOR ADVANCED STUDY

ANNALS OF MATHEMATICS
Vol. 45, No. 4, October, 1944

GROUP INVARIANCE OF CAUCHY'S FORMULA IN SEVERAL VARIABLES

BY S. BOCHNER

(Received February 1, 1944)

1. If we iterate Cauchy's formula

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f(\zeta)\,d\zeta}{\zeta - z} \tag{1}$$

we obtain

$$f(z_1, \cdots, z_k) = \frac{1}{2\pi i}\int_{C_k} f(\zeta_1, \cdots, \zeta_k)\,\frac{d\zeta_1 \cdots d\zeta_k}{(\zeta_1 - z_1)\cdots(\zeta_k - z_k)}. \tag{2}$$

In (2), C_k is a k-dimensional "characteristic manifold" of a special topological type, its prototype being the torus

$$|\zeta_1| = 1, \cdots, |\zeta_k| = 1. \tag{3}$$

The special nature of this manifold corresponds to the special nature of the kernel

$$G(\zeta; z) = \frac{1}{(\zeta_1 - z_1)\cdots(\zeta_k - z_k)} \tag{4}$$

under the integral, and S. Bergmann[1] and A. Weil[2] have studied alterations of the formula to be brought about by suitably modifying both C_k and $G(\zeta; z)$.

In the present paper we will deal with the same question but from a different approach. In the work of Bergmann and Weil the emphasis is put on obtaining a formula to suit a prescribed domain of validity in z, whereas our emphasis will be on retaining prescribed structural features of the formula as such. We note that in the special case (4), the kernel depends only on the differences $\zeta_1 - z_1, \cdots, \zeta_k - z_k$, and this feature will be retained unaltered. We further note that the torus (3) is invariant and transitive under the group of transformations

$$\zeta_1 \to e^{i\alpha_1}\zeta_1, \cdots, \zeta_k \to e^{i\alpha_k}\zeta_k.$$

This feature will be likewise retained but other groups of affine transformations will be admitted.

Part I will deal with the group of translations, Part II with some cases of the unitary group, and Part III with general groups. The general case will be based

[1] *On the surface integrals of functions of two complex variables*, Amer. Jour. of Math. 63 (1941), 295–318.

[2] *L'integrale de Cauchy et les fonctions de plusieurs variables*, Math. Annalen, 111 (1935), 178–182.

on an operational version of Cauchy's theorem in which no reference is made to "realization" by integrals.

PART I

OCTANT-SHAPED DOMAINS

2. Tubes

Given any k complex variables

$$z_p = x_p + iy_p\,; \qquad p = 1, \cdots, k, \tag{5}$$

we consider a k-dimensional domain P in the space of the real variables $x = (x_1, \cdots, x_k)$, and with the point set P we construct in the $(2k)$-dimensional space of the variables (5) the domain

$$\{x \in P;\ -\infty < y_p < \infty,\ p = 1, \cdots, k\}. \tag{6}$$

Such a domain has been called a *tube*,[3] and the point set P its *base*. A tube will be usually denoted by the letter T. We will say that the domain P is *radiated* if it is convex, does not contain the origin $x = (0, \cdots, 0)$ but is composed of rays which emanate from the origin. That is, if $x^0 = (x_1^0, \cdots, x_k^0)$ is a point of P, then the segment

$$x = tx^0, \qquad 0 < t < \infty \tag{7}$$

also belongs to P. We now pick another set of k real variables denoting them by $\alpha = (\alpha_1, \cdots, \alpha_k)$ and we consider all those α for which the inequality

$$\alpha_1 x_1 + \cdots + \alpha_k x_k > 0 \tag{8}$$

holds for all $x \in P$. They obviously form a convex set. We make the explicit assumption that the set is not empty and that it actually has an interior thus being the closure of a domain. This (open) domain will be denoted by G and called the conjugate to P. The domain G is again radiated. We now set up the integral

$$\int_G e^{-(\alpha_1 x_1 + \cdots + \alpha_k x_k)}\, dv_\alpha \tag{9}$$

where $dv_\alpha = d\alpha_1 \cdots d\alpha_k$ and we make the final explicit assumption that this integral is boundedly convergent in every compact subset of P. If all these assumptions are fulfilled we will say that the tube T is *octant-shaped*, and we associate with it the "kernel"

$$K(z_1, \cdots, z_k) = \int_G e^{-(z_1\alpha_1 + \cdots + z_k\alpha_k)}\, dv_\alpha \tag{10}$$

which is obviously defined and analytic in all of T.

---

[3] S. BOCHNER, *Bounded analytic functions in several variables and multiple Laplace integrals*, Amer. Jour. of Math. 59 (1937), 732–738.

The boundary of T is likewise "tubular", and the most important part (for us) of the boundary will be the pointset

$$\{\zeta_p = 0 + i\eta_p;\ -\infty < \eta_p < \infty;\ p = 1, \cdots, k\}. \tag{11}$$

It will be called the *spine* of T. The Euclidean volume element on the spine is, of course, $dv\eta = d\eta_1 \cdots d\eta_k$; however there will be a formal advantage in using not the volume element as such but the "oriented" external differential

$$d\zeta_1 \cdots d\zeta_k \equiv i^k d\eta_1 \cdots d\eta_k \equiv i^k dv\eta, \tag{12}$$

and whenever this differential will occur under the integral, the spine as a manifold will be denoted by iE_k.

3. Functions of integrable square[3]

Functions in tubes can be subjected to Fourier analysis. If $f(z)$ is analytic in T then we say that it is of integrable square, or belongs to L_2, if there exists a finite constant $M = M_f$, such that

$$\frac{1}{(2\pi)^k}\int_{E_k} |f(x + iy)|^2\, dv_y \leqq M_f \tag{13}$$

for all x in P (and not only for every compact subset of P). If (13) holds then $f(x + iy)$, for each x, can be represented as a Fourier integral in y, and these representations coalesce into an absolutely convergent Laplace-integral

$$f(z) = \int_{E_k} e^{-(z_1\alpha_1 + \cdots + z_k\alpha_k)} \varphi(\alpha_1, \cdots, \alpha_k)\, dv_\alpha . \tag{14}$$

The function $\varphi(\alpha)$ is measurable, and by the equality of Parseval-Plancherel we have

$$\int_{E_k} |\varphi(\alpha)|^2 e^{-2(x_1\alpha_1 + \cdots)}\, dv_\alpha = \frac{1}{(2\pi)^k}\int_{E_k} |f(x + iy)|^2\, dv_y ; \tag{15}$$

therefore

$$\int_{E_k} |\varphi(\alpha)|^2 e^{-2(x_1\alpha_1 + \cdots)}\, dv_\alpha \leqq M_f . \tag{16}$$

If we now recall the definition of the conjugate set G, and if we realize that the Lebesgue measure of the boundary of G is zero (since G is convex), then a simple argument will show that the function $\varphi(\alpha)$ must be zero, almost everywhere, outside the conjugate domain G. Thus, for instance, we can rewrite relations (14) and (16) into

$$f(z) = \int_G e^{-(z_1\alpha_1 + \cdots)} \varphi(\alpha)\, dv_\alpha \tag{17}$$

and

(18) $$\int_G |\varphi(\alpha)|^2 e^{-2(x_1\alpha_1+\cdots)}\, dv_\alpha \leqq M_f.$$

Since $x = (0, \cdots, 0)$ is a boundary point of P and for all points of P relation (8) holds, we may now derive from (18) the limit relations

(19) $$\int_G |\varphi(\alpha)|^2\, dv_\alpha \leqq M_f$$

and

(20) $$\lim_{x\to 0} \int_G |\varphi(\alpha)|^2(1 - e^{-2(x_1\alpha_1+\cdots)})\, dv_\alpha = 0.$$

The first implies the existence of a function of integrable square

(21) $$f(i\eta) \sim \int_G e^{-i(\eta_1\alpha_1+\cdots)}\varphi(\alpha)\, dv_\alpha,$$

and the second establishes the limit relation

(22) $$\lim_{x\to 0} \int_{-\infty}^{\infty} |f(i\eta) - f(x + i\eta)|^2\, dv_\eta = 0.$$

Thus our function $f(z)$ has automatically generalized boundary values (in the strong topology of L_2) almost everywhere on the spine of T; also, these boundary values are identical with the limit-values

$$\lim_{x\to 0} f(x + i\eta)$$

whenever such limit values exist on a measurable set of the spine.

Now, for each z in T, $K(z - i\eta)$ is a function of integrable square in η, namely

(23) $$K(z - i\eta) \sim \int_G e^{i(\eta_1\alpha_1+\cdots)} e^{-(z_1\alpha_1+\cdots)}\, dv_\alpha.$$

By the theory of Fourier convolutions (resultants), relations (21) and (23) imply the relation

(24) $$\frac{1}{(2\pi)^k}\int_{E_k} f(i\eta)K(z - i\eta)\, dv_\eta = \int_G e^{-(z_1\alpha_1+\cdots)}\varphi(\alpha)\, dv_\alpha.$$

If we combine this with (17) and recall (12), we finally obtain the following theorem.

THEOREM 1. *If T is an octant-shaped tube and $K(z)$ is its kernel as defined by* (10), *then for functions $f(z)$ of class L_2 we have the formula*

(25) $$f(z) = \frac{1}{(2\pi i)^k}\int_{iE_k} K(z - \zeta)f(\zeta)\, d\zeta_1 \cdots d\zeta_k$$

for all z in T.

The most obvious case of an octant-shaped domain is the tube

$$x_1 > 0, \cdots, x_k > 0.$$

Its basis is "self-conjugate" and the kernal is $1/z_1 \cdots z_k$. However, a more significant type of domain, also self-conjugate, has come to the fore recently and we will discuss it in some detail.

4. Matrices[4]

If $k = n^2$, then the variables

$$z_1, \cdots, z_k \tag{26}$$

can be re-indexed into a square scheme

$$z_{pq}, \qquad p, q = 1, \cdots, n. \tag{27}$$

The space of these independent complex variables will be called the *non-symmetric (matrix) space.* The letter z will sometimes denote the set of variables in either indexing, but most of the time it will denote the matrix (27) as an algebraic concept.

If the matrix is symmetric, $z_{pq} = z_{qp}$, then we choose the elements

$$\begin{matrix} z_{11}, & z_{12}, & \cdots, & z_{1n} \\ & z_{22}, & \cdots, & z_{2n} \\ & & & \cdot \\ & & & \cdot \\ & & & z_{nn} \end{matrix} \tag{28}$$

as independent complex variables, and the letter z will again denote either the set of these variables or the resulting matrix. In this case $k = n(n + 1)/2$, and the space will be called the symmetric space.

We will require a lemma.

LEMMA 1. *In non-symmetric space, if a and b are arbitrary constant matrices, then the Jacobian*

$$J = \frac{\partial(w_1, \cdots, w_k)}{\partial(z_1, \cdots, z_k)}$$

of the linear transformation

$$w = bza \tag{29}$$

has the value

$$J = (\det | ba |)^n. \tag{30}$$

In symmetric space, if a is an arbitrary matrix and a' is its transposed, then the Jacobian of the transformation

---

[4] See C. L. SIEGEL, *Symplectic geometry*, Amer. Jour. of Math. 65 (1943), 8–10.

$$w = a'za \tag{31}$$

has the value

$$(\det | a |)^{n+1}. \tag{32}$$

PROOF. Since the Jacobian of successive transformations is multiplicative, it is sufficient to verify (30) for the special cases: $a = 1$ and b arbitrary, or $b = 1$ and a arbitrary, and we will only verify the second case. Thus we have to prove that

$$J(a) = (\det | a |)^n \tag{33}$$

for the transformation $w = za$. In the first place, (33) can be verified by direct computation for a diagonal matrix a. In the second place, we have $J(b^{-1}cb) = J(b^{-1})J(c)J(b) = J(c)$, and thirdly we have $\det | b^{-1}cb | = \det | c |$. Thus, (33) holds for any a of the form $b^{-1}cb$ with c diagonal. The general case follows now by analytic continuation in the elements of the matrix a.

In the symmetric case we have to assume $b = a'$ in order to make sure that a symmetric z will transform into a symmetric w. The proof proceeds along the same lines as before.

5. The symmetric case

We write $z = x + iy$, where $x = \{x_{pq}\}$ is the matrix of the real parts and $y = \{y_{pq}\}$ is the matrix of the imaginary parts of the variables (28). The matrices x and y are real and symmetric, and hence hermitian. We now consider the set of those matrices x whose roots are all strictly positive, in symbols $x > 0$. The corresponding pointset will hence forth be denoted by P_n^0 and the tube with P_n^0 as base will be denoted by T_n^0. It is easily seen that P_n^0 is a radiated set. We now take another real symmetric matrix $\beta = \{\beta_{pq}\}$ in which again $(n(n+1))/2$ elements are variable, and we set up the familiar expression

$$\operatorname{tr}(x\beta) = \sum_{p,q} x_{pq}\beta_{qp}. \tag{34}$$

In the double sum each non-diagonal term occurs twice, and it can be written as $x_1\alpha_1 + \cdots + x_k\alpha_k$ where each α_j is a β_{pq} or double a β_{pq}; also

$$dv_\alpha = 2^{n(n-1)/2} dv_\beta. \tag{35}$$

Relation (8) states that $\operatorname{tr}(x\beta) > 0$; and the latter will hold for all $x > 0$ if and only if $\beta > 0$ (as a matrix). Thus the conjugate domain G consist again of all positive matrices, and we will simply denote it likewise by P_n^0. For the kernel, if any, we now have the equation

$$K(x) = 2^{n(n-1)/2} \int_{P_n^0} e^{-\operatorname{tr}(x\beta)} dv_\beta. \tag{36}$$

This integral occurs in the work of Siegel.[5] He actually considers the more general integral

[5] C. L. SIEGEL, *Ueber die analytische Theorie der quadratischen Formen*, Annals of Math. 36 (1935), 585, Hilfsatz 37.

$$\int_{P_n^0} e^{-\operatorname{tr}(x\beta)} (\det |\beta|)^{s-(n+1)/2} \, dv_\beta \tag{37}$$

and he shows that it has the value

$$(\det |x|)^{-s} \prod_{m=1}^{n} \pi^{(m-1)/2} \Gamma\left(s - \frac{m-1}{2}\right). \tag{38}$$

(We will indicate the process of evaluation in the discussion of the non-symmetric case.) If we now put $s = (n+1)/2$, then Theorem 1 implies the following theorem:

THEOREM 2. *In the symmetric case we have the formula*

$$f(z) = c_n^0 \int_{iB_k^0} \frac{f(\zeta) \, d\zeta_1 \cdots d\zeta_k}{(\det |z_{pq} - \zeta_{pq}|)^{(n+1)/2}} \tag{39}$$

where

$$c_n^0 = \frac{\Gamma\left(\frac{2}{2}\right) \Gamma\left(\frac{3}{2}\right) \cdots \Gamma\left(\frac{n+1}{2}\right)}{2^n \pi^{(n^2+3n)/2} i^{n(n+1)/2}} \tag{40}$$

for functions of integrable square in T_n^0.

We remark that for even n, the value of the denominator

$$(\det |z_{pq} - \zeta_{pq}|)^{(n+1)/2}$$

is to be obtained by analytic continuation, which is easily proven to be unique, from its initial value $+1$, for $z = 1$ (matrix), $\zeta = 0$ (matrix).

6. The non-symmetric case

The analysis is now more involved than in the symmetric case, however the final formula will be even simpler in structure. We will employ two systems of variables which we will denote by $\{z_{pq}\}$ and $\{Z_{pq}\}$ respectively. They shall be connected by the affine transformation

$$Z_{pp} = z_{pp}; \quad p = 1, \cdots, k,$$

$$\left.\begin{aligned} Z_{pq} &= \frac{z_{pq} + z_{qp}}{2}, & z_{pq} &= Z_{pq} + iZ_{qp} \\ Z_{qp} &= \frac{z_{pq} - z_{qp}}{2i}, & z_{qp} &= Z_{pq} - iZ_{qp} \end{aligned}\right\}, \qquad p < q.$$

If we introduce the real and imaginary parts of all variables, $z_{pq} = x_{pq} + iy_{pq}$, $Z_{pq} = X_{pq} + iY_{pq}$, then we have the relations

$$X_{pp} = x_{pp} \qquad Y_{pp} = y_{pp}$$

$$X_{pq} = \frac{x_{pq} + x_{qp}}{2} \qquad Y_{pq} = \frac{y_{pq} + y_{qp}}{2}$$

$$X_{qp} = \frac{y_{pq} - y_{qp}}{2} \qquad Y_{qp} = \frac{x_{qp} - x_{pq}}{2}$$

where again $p < q$. The coordinates z_{pq} will be called *primary* coordinates, and the coordinates Z_{pq} will be called *secondary* coordinates, and we will employ both systems concurrently.

We introduce a third system of quantities in terms of the primary coordinates. If z is a matrix and if z^* is the conjugate-complex transposed matrix ("adjoint" in Hilbert Space terminology), then we put

$$r = \frac{z + z^*}{2}.$$

Hence

$$r_{pq} = \frac{z_{pq} + \bar{z}_{qp}}{2}$$

or

$$\begin{aligned} r_{pp} &= X_{pp} \\ \left.\begin{aligned} r_{pq} &= X_{pq} + i\,X_{qp} \\ r_{qp} &= X_{pq} - i\,X_{qp} \end{aligned}\right\}&, \quad p < q. \end{aligned}$$

The matrix r is hermitian. Furthermore if

(41) $$\varphi(r_{11}, r_{12}, r_{21}, \cdots)$$

is an analytic expression in the quantities r_{pq}; if we view the quantities $\{r_{pq}\}$ as dependent on $\{X_{pq}\}$; and if we then perform the analytic continuation of replacing each X_{pq} by Z_{pq}; then the function (41) will go over into

$$\varphi(Z_{11}, Z_{12} + iZ_{21}, Z_{12} - iZ_{21}, \cdots)$$

which in its turn is

(42) $$\varphi(z_{11}, z_{12}, z_{21}, \cdots).$$

In other words, if we consider a domain which is a tube in terms of secondary coordinates, then the transition from (41) to (42) is the continuation of the function φ from the base of the tube into the tube proper.

Such a tube will be now defined. It shall consist of those points of the space for which all roots of the matrix r are strictly positive, $r > 0$. Its basis (in terms of the components X_{pq}) will be denoted by P_n, and the resulting tube, which can also be described as

$$r \in P_n; \qquad -\infty < Y_{pq} < \infty,$$

will be denoted by T_n. It is easily seen that P_n is radiated, and the computations to follow will imply that T_n is octant-shaped. In order to describe the conjugate set G we have to consider the expression

(43) $$\operatorname{tr}(X\beta) \equiv \sum_{pq} X_{pq}\beta_{pq}$$

for an arbitrary real matrix β. If we put

$$\left.\begin{aligned} \alpha_{pp} &= \beta_{pp} \\ \alpha_{pq} &= \frac{\beta_{pq} + i\beta_{qp}}{2} \\ \alpha_{qp} &= \frac{\beta_{pq} - i\beta_{qp}}{2} \end{aligned}\right\}, \qquad p < q. \tag{44}$$

then (43) is identical with

$$\operatorname{tr}(r\alpha) = \sum_{p,q} r_{pq}\alpha_{qp}. \tag{45}$$

For fixed α, if α is diagonal, then $\operatorname{tr}(r\alpha)$ will be positive for all positive r, if and only if, α itself is positive. For α non-diagonal we can put $\alpha = u\alpha^0u^{-1}$ where α^0 is diagonal and u is unitary, and the relations

$$\operatorname{tr}(r\alpha) = \operatorname{tr}(ru\alpha^0u^{-1}) = \operatorname{tr}([u^{-1}ru]\alpha^0)$$

will lead to the same conclusion. Thus we see that G is a domain in the β-variables which in terms of the auxiliary hermitian matrix α is again described by $\alpha > 0$. This domain will be denoted specifically by G_n.

As for the kernel of T_n, its value in P_n is given by the integral

$$K(r) = \int_{G_n} e^{-\operatorname{tr}(r\alpha)}\, dv_\beta \tag{46}$$

and, by our previous remark, its value in primary coordinates of T_n will be $K(z)$. Suppose first that r is diagonal, $r_{pp} = \lambda_p$; $r_{pq} = 0$, $p \neq q$. The transformation

$$\alpha_{pq} \to \frac{\alpha_{pq}}{\sqrt{\lambda_p\lambda_q}}, \qquad p \lesseqgtr q,$$

or rather the corresponding transformation in the β-variables, transforms G_n into itself, and replaces dv_β by $(dv_\beta)/((\lambda_1 \cdots \lambda_n)^n.)$ Since $\lambda_1 \cdots \lambda_n = \det|r|$ and $\alpha_{pp} = \beta_{pp}$, we therefore obtain, for r diagonal,

$$K(r) = \frac{a_n}{(\det|r|)^n}, \tag{47}$$

where

$$a_n = \int_{G_n} e^{-\operatorname{tr}(\beta)}\, dv_\beta. \tag{48}$$

If $r = ur^0u^{-1}$, where r^0 is diagonal and u is unitary, we replace in the integral, α by $u\alpha u^{-1}$. On the basis of Lemma 1 in connection with the identity $\operatorname{tr}\{ur^0u^{-1}\alpha\} = \operatorname{tr}\{r^0u^{-1}\alpha u\}$ it is now possible to verify that $K(r) = K(r^0)$. Since we also have $\det|r| = \det|r^0|$, we may thus conclude that (47) holds for an arbitrary hermitian matrix r. In order to compute the constant a_n, it is advisable to set up the more general expression

$$A_n(s) = \int_{G_n} e^{-\mathrm{tr}(\beta)} (\det |\alpha_{pq}|)^{s-n} \, dv_\beta, \tag{49}$$

for $s \geqq n$, and to prove the inductive relation

$$A_n(s) = (4\pi)^{n-1} \Gamma(s - n + 1) A_{n-1}(s) \tag{50}$$

for $n \geqq 2$. The coordinates

$$\beta_{\rho\sigma}; \quad \rho, \sigma = 2, 3, \cdots, n \tag{51}$$

are subject to the condition $\det | \alpha_{\rho\sigma} | > 0$, and their totality describes a point-set that may be denoted by G_{n-1}. If the quantities (51) are fixed, then the remaining coordinates

$$\beta_{11}; \quad \beta_{1\rho}, \beta_{\sigma 1}; \quad \rho, \sigma = 2, \cdots, n \tag{52}$$

are variables subject to the restriction

$$\det \begin{vmatrix} \beta_{11}, & \dfrac{\beta_{1\rho} + i\beta_{\sigma 1}}{2} \\ \dfrac{\beta_{1\rho} - i\beta_{\sigma 1}}{2}, & \alpha_{\rho\sigma} \end{vmatrix} > 0, \tag{53}$$

and we are going to integrate (49) with respect to the variables (52) first. In order to simplify the computation we will first assume that the matrix $| \alpha_{\rho\sigma} |$ is a diagonal matrix $\begin{vmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_n \end{vmatrix}$ and the value of its determinant will be denoted by D. Since $tr(\beta) = \beta_{11} + \beta_{22} + \cdots$ we thus have to compute the integral

$$D^{s-n} \int e^{-\beta_{11}} \left(\beta_{11} - \sum_{\rho=2}^{n} \frac{\beta_{1\rho}^2 + \beta_{\rho 1}^2}{4\lambda_\rho} \right)^{s-n} d\beta_{11} \prod_{\rho=2}^{n} d\beta_{1\rho} \, d\beta_{\rho 1}, \tag{54}$$

the integration extending over the point set

$$\beta_{11} > 0, \quad \beta_{11} - \sum_{\rho=2}^{n} \frac{\beta_{1\rho}^2 + \beta_{\rho 1}^2}{4\lambda_\rho} > 0. \tag{55}$$

In terms of new coordinates

$$\beta = \beta_{11}, \quad \xi_{\rho-1} = \frac{\beta_{1\rho}}{\sqrt{4\lambda_\rho \beta_{11}}}, \quad \eta_{\rho-1} = \frac{\beta_{\rho 1}}{\sqrt{4\lambda_\rho \beta_{11}}}$$

the integral is

$$4^{n-1} D^{s-n+1} \int e^{-\beta} \beta^{s-1} \left(1 - \sum_{\rho=1}^{n-1} \xi_\rho^2 + \eta_\rho^2 \right)^{s-n} d\beta \prod d\xi_\rho \, d\eta_\rho$$

and the point set is

$$\beta > 0, \quad 1 - \sum_{\rho=1}^{n-1} \xi_\rho^2 + \eta_\rho^2 > 0.$$

A direct computation gives the value

$$D^{s-n+1}(4\pi)^{n-1}\Gamma(s - n + 1),$$

and a re-check will show that we obtain the same value for non-diagonal matrices $| \alpha_{pr} |$. This completes the proof of (50). Since $A_1(s) = \Gamma(s)$, we finally obtain

$$A_n(s) = (4\pi)^{n(n-1)/2} \prod_{m=1}^{n} \Gamma(s - m + 1) \tag{56}$$

and in particular

$$a_n = (4\pi)^{n(n-1)/2} \prod_{m=1}^{n-1} m! \tag{57}$$

If we insert (47) in (25) we obtain the integral

$$\frac{a_n}{(2\pi i)^{n^2}} \int_{iB_k} \frac{f(\zeta)\, dZ_1^0 \cdots dZ_k^0}{(\det | z_{pq} - \zeta_{pq} |)^n},$$

where ζ_{pq} are primary and Z_p^0 are secondary coordinates on the spine. Finally, if we replace secondary coordinates by primary coordinates even on the spine, we obtain the following theorem.

THEOREM 3. *In the non-symmetric case we have the formula*

$$f(z) = \frac{1!2! \cdots (n-1)!}{(2\pi i)^{n(n+1)/2}} \int \frac{f(\zeta)\, d\zeta_1 \cdots d\zeta_k}{(\det | z_{pq} - \zeta_{pq} |)^n} \tag{58}$$

for functions in T_n which are of integrable square relative to secondary coordinates; the integration extending over the spine of T_n.

In Part II we will transfer formulas (39) and (58) from the tube to a type of unit sphere, thus bringing them much closer to being Cauchy formulas, and in Part III we will make an analogous investigation for the general formula (25).

PART II

SPHERE-SHAPED DOMAINS

7. The Cayley transformation

If z is any matrix and z^* is the adjoint matrix, then the products zz^* and z^*z are non-negative hermitian and, although not necessarily equal, they have the same roots.[6] The set of points z for which $zz^* < 1$, (or $z^*z < 1$) will be called the "unit sphere". It will be denoted by S_n in the non-symmetric, and by S_n^0 in the symmetric case. Thus, explicitly, $\{z_{pq}\}$ belongs to S_n (or S_n^0) if for every vector $(\lambda_1, \cdots, \lambda_n) \neq (0, \cdots, 0)$,

$$z_{pr}\bar{z}_{qr}\lambda_p\bar{\lambda}_q < \lambda_r\bar{\lambda}_r$$

or if

$$z_{rp}\bar{z}_{rq}\lambda_p\bar{\lambda}_q < \lambda_r\bar{\lambda}_r .$$

[6] Compare P. R. HALMOS, *Finite dimensional vector spaces*, Princeton (1942) 138–140.

Unit-matrices ζ are those for which $\zeta\zeta^* = 1$. Thus if U_n (or U_n^0 in the symmetric case) is the set of all unitary matrices, then U_n is a proper part of the boundary of S_n.

We now claim that for $\zeta \in U_n$, $z \in S_n$ we have

$$\det | \zeta - z | \neq 0. \tag{59}$$

Since $\zeta - z = (1 - z\zeta^*)\zeta$ and $z\zeta^*(z\zeta^*)^* = z\zeta^*\zeta z^* < 1$ it is obviously sufficient to show that $\det | 1 - z | \neq 0$ for $z \in S_n$. If this were false, there would exist a vector $\lambda \neq 0$ for which $\lambda_r = z_{rp}\lambda_p$. On multiplying this relation with its conjugate complex, we would obtain $\lambda_r\bar{\lambda}_r = \overline{z_{rp}z_{rq}}\lambda_p\bar{\lambda}_q$ which is impossible for z in S_n. Since $\zeta = -1$ is unitary, we see that $\det | 1 + z | \neq 0$, and thus for $z \in S_n$ we can set up the quotient

$$w = \frac{1 - z}{1 + z} \tag{60}$$

("Cayley transformation"). Obviously

$$\begin{aligned} w + w^* &= (1 + z)^{-1}(1 - z) + (1 - z^*)(1 + z^*)^{-1} \\ &= (1 + z)^{-1}\{(1 - z)(1 + z^*) + (1 + z)(1 - z^*)\}(1 + z^*)^{-1} \\ &= 2(1 + z)^{-1}(1 - zz^*)(1 + z^*)^{-1}. \end{aligned}$$

Now, in general, if p is positive hermitian and q is arbitrary, then qpq^* is again positive hermitian. Thus, $1 - zz^* > 0$ implies $w + w^* > 0$, and the latter inequality characterizes a point of the tube T_n in primary coordinates.

Conversely, let $w \in T_n$. We certainly have $\det | 1 + w | \neq 0$. Otherwise we would have a relation $\lambda_r = - w_{rp}\lambda_p$ which in its turn would lead to

$$2\lambda_r\bar{\lambda}_r = - (w_{pq} + \bar{w}_{qp})\lambda_p\bar{\lambda}_q .$$

This however contradicts $w + w^* > 0$. Thus we can set up the quotient

$$z = \frac{1 - w}{1 + w} \tag{61}$$

which is the formal converse to (60), and we obtain

$$\begin{aligned} 1 - zz^* &= 1 - (1 + w)^{-1}(1 - w)(1 - w^*)(1 + w^*)^{-1} \\ &= (1 + w)^{-1}\{(1 + w)(1 + w^*) - (1 - w)(1 - w^*)\}(1 + w^*)^{-1} \\ &= 2(1 + w)^{-1}\{w + w^*\}(1 + w^*)^{-1}. \end{aligned}$$

This implies $1 - zz^* > 0$. Hence we have proven

LEMMA 2. *The Cayley transformation is a one-to-one map between S_n and T_n; or between S_n^0 and T_n^0.*

Now, if ζ is any point of U_n, and if $\det | 1 + \zeta | \neq 0$, then

$$\omega = \frac{1 - \zeta}{1 + \zeta} \tag{62}$$

is a point of the spine iE_k. Conversely, if ω is any point whatsoever of iE_k, then $\zeta = (1-\omega)/(1+\omega)$ is a point of U_n. Thus, if we denote by A_1 the subset of U_n which is characterized by $\det|1+\zeta| = 0$, then (62) is a one-to-one map between $V_1 = U_n - A_1$ and iE_k. More generally, take any fixed number ρ of absolute value 1, denote by A_ρ the subset of U_n for which $\det|1+\rho\zeta| = 0$, and put $V_\rho = U_n - A_\rho$ then the transformation

$$\rho\zeta = \frac{1-\omega}{1+\omega}$$

is a one-to-one map between V_ρ and iE_k. Now, if $\rho_0 = 1, \rho_1, \cdots, \rho_n$ are $n+1$ different values of ρ then $\{V_{\rho_0}, V_{\rho_1}, \cdots, V_{\rho_n}\}$ is a complete covering of U_n, since the polynomial $\det|1+\rho\zeta|$ cannot have more than $n+1$ roots ρ. Finally, the intersection of A_1 with any V_ρ is the image in U_n of the pointset

$$\det\left|1+\frac{1}{\rho}\,\frac{1-\omega}{1+\omega}\right| = 0$$

in iE_k, and this is an analytic manifold of dimension $\leqq k-1$. Thus, U_n (and similarly U_n^0) is a k-dimensional analytic manifold and we have

LEMMA 3. *The Cayley transformation is a one-to-one map between $U_n - A$ and iE_k, where A is a lower-dimensional analytical subset of U_n; similarly, between $U_n^0 - A^0$ and iE_k^0.*

Finally, we will state a last pre-requisite.

LEMMA 4. *The Jacobian $\partial(z)/\partial(w)$ of the transformation* (61) *is*

$$(-2)^{n^2}\frac{1}{(\det|1+w|)^{2n}}$$

in the non-symmetric case, and

$$(-2)^{n(n+1)/2}\frac{1}{(\det|1+w|)^{n+1}}$$

in the symmetric case.

In fact, from $z(1+w) = 1-w$ we obtain successively

$$\begin{aligned} dz\cdot(1+w) + z\cdot dw &= -\,dw \\ dz\cdot(1+w) &= -(1+z)\cdot dw \\ dz &= -(1+z)\cdot dw\cdot(1+w)^{-1} \\ &= -2(1+w)^{-1}\cdot dw\cdot(1+w)^{-1} \end{aligned}$$

and Lemma 4 follows now from Lemma 1.

8. The non-symmetric case

We note that U_n, being a group space, is orientable. Therefore, the differential $d\zeta_1 \cdots d\zeta_k$ can be defined in the large, and its algebraic sign can be changed by a change of orientation.

THEOREM 4. *If $f(z)$ is analytic and bounded in S_n and has boundary values in the ordinary sense at all points of U_n, then the formula*

$$f(z) = \frac{1!2! \cdots (n-1)!}{(2\pi i)^{n(n+1)/2}} \int_{U_n} \frac{f(\zeta)\, d\zeta_1 \cdots d\zeta_k}{\det |\zeta_{pq} - z_{pq}|)^n} \tag{63}$$

is valid for z in S_n.

If we apply the transformations (61) and (62), we have by Lemma 4,

$$d\zeta_1 \cdots d\zeta_k = \pm 2^{n^2} \frac{d\omega_1 \cdots d\omega_k}{(\det |1 + \omega|)^{2n}}$$

and the relation

$$\begin{aligned} \zeta - z &= (1 + \omega)^{-1}(1 + \omega) - (1 - w)(1 + w)^{-1} \\ &= 2(1 + \omega)^{-1}[w - \omega](1 + w)^{-1} \end{aligned}$$

implies the equality

$$(\det |\zeta - z|)^n = 2^{n^2}(\det |w - \omega|)^n(\det |(1 + \omega)(1 + w)|)^{-n}.$$

Also the exceptional subset A of U_n, which is not affected by the transformation, can be neglected in the integral (63) on the basis of Lemma 3. Thus, formula (63), whether true or false, is equivalent with formula

$$\mathrm{F}(w) = c_n \int_{iE_k} \frac{\mathrm{F}(\omega)\, d\omega_1 \cdots d\omega_k}{(\det |w - \omega|)^n} \tag{64}$$

where

$$F(w) = \frac{f(z(w))}{(\det |1 + w|)^n} \tag{65}$$

and

$$c_n = \frac{1!2! \cdots (n-1)!}{(2\pi i)^{n(n+1)/2}}. \tag{66}$$

By assumption, $f(z(w))$ is bounded; therefore $F(w)$ is of integrable square in T_n, and its boundary values on the spine derive from the values $f(\zeta)$ on U_n. Thus (64) is true by Theorem 3, and the proof of Theorem 4 is completed.

9. Group-invariant volume

We note that in non-symmetric space, by Lemma 1, the differential form

$$c_n \frac{d\zeta_1 \cdots d\zeta_k}{(\det |\zeta_{pq}|)^n} \tag{67}$$

is invariant for arbitrary transformations $\zeta \to a\zeta b$. For σ unitary, $\zeta \to \sigma\zeta$ is an automorphism of U_n as group space and we can now deduce that (67) is the

volume element of the group invariant volume on U_n. In fact, we can define the additive set function

$$\mu(B) = c_n \int_B \frac{d\zeta_1 \cdots d\zeta_k}{(\det | \zeta |)^n}$$

for Borel sets on U_n. It is group invariant, and it is absolutely continuous relative to the group volume, therefore by a general theorem[7] it is a constant multiple of the volume. Also $\mu(U_n) = 1$, and thus it is identical with the volume. We will now denote (67) by dv, and since $\zeta - z = (1 - z\zeta^*)\zeta$ we now obtain

Theorem 5. *Under the assumptions of Theorem 4 we have*

$$f(z) = \int_{U_n} \frac{f(\zeta)\, dv}{(\det | 1 - z\zeta^* |)^n} \tag{68}$$

where dv is the group invariant volume element of U_n, with the entire group having volume 1.

If we introduce the kernel

$$H(z; \bar{\zeta}) = (\det | 1 - z\zeta^* |)^{-n} \tag{69}$$

then formula (68) reads

$$f(z) = \int H(z; \bar{\zeta}) f(\zeta)\, dv. \tag{70}$$

The kernel (69) is analytic in $z_1, \cdots, z_k$; $\bar{\zeta}_1, \cdots, \bar{\zeta}_k$ and has a power series expansion

$$H(z; \bar{\zeta}) = \sum_{p=0}^{\infty} H_p(z; \bar{\zeta}) \tag{71}$$

where each $H_p(z; \bar{\zeta})$ is a homogeneous polynomial (with real coefficients) of dimension p in either z or $\bar{\zeta}$. Also $H(z; \bar{\zeta}) = H(\bar{\zeta}; z)$ and therefore

$$H_p(z; \bar{\zeta}) = H_p(\bar{\zeta}; z). \tag{72}$$

It is easily seen that the series (71) is uniformly convergent for ζ varying in the closure of S_n and z varying in some neighborhood of the origin; (actually, z may vary in any compact subset of S_n). On substituting (71) in (70) we therefore obtain

$$f(z) = \sum_{p=0}^{\infty} f_p(z) \tag{73}$$

where

$$f_p(z) = \int H_p(z; \bar{\zeta}) f(\zeta)\, dv. \tag{74}$$

[7] S. Bochner, *Additive set functions on groups*, Annals of Math. 40 (1939), 787, Theorem 3.

By (74), $f_p(z)$ must be of dimension p or identically 0, therefore the series (73) is the diagonal form of the power series of $f(z)$. From this we conclude that

$$\int H_p(z;\bar{\zeta})f_q(\zeta)\,dv = \begin{cases}0, & p \neq q\\ f_p(z), & p = q.\end{cases} \tag{75}$$

If we multiply both sides by $\overline{g_p(z)}$ and recall (72), and integrate with respect to z, then we obtain the following theorem on unitary groups.

THEOREM 6. *If $g_p(\zeta)$ and $f_q(\zeta)$ are homogeneous polynomials of different dimensions then*

$$\int_{U_n} g_p(\zeta)\,\overline{f_q(\zeta)}\,dv = 0. \tag{76}$$

Also, if $\{\varphi_r(\zeta)\}$ is a complex system of orthonormal polynomials then

$$H(z;\bar{\zeta}) = \sum_{r=0}^{\infty} \varphi_r(z)\overline{\varphi_r(\zeta)}, \tag{77}$$

the latter series being summed diagonally.

As for series (77), since $\{\varphi_r(\zeta)\}$ is a complete system of homogeneous polynomials, there exists an expansion, diagonally convergent, of the form

$$H(z;\bar{\zeta}) = \sum c_r(z)\overline{\varphi_r(\zeta)}$$

with suitable polynomials $c_r(z)$. Their identity with $\varphi_r(z)$ follows now from (70) if applied to $f(z) = \varphi_r(z)$.

10. The symmetric case

This case has a complication. We do not know whether the manifold U_n^0, which is not a group space is orientable, and therefore we are not justified in setting up the differential $d\zeta_1 \cdots d\zeta_k$ for the entire manifold. Nevertheless we can set it up for the manifold $V_n^0 = U_n^0 - A^0$ where A^0 is the exceptional subset occurring in Lemma 3. Thus, in analogy to (63), we can now obtain the formula

$$f(z) = c_n^0 \int_{V_n^0} \frac{f(\zeta)\,d\zeta_1 \cdots d\zeta_k}{(\det|\zeta_{pq} - z_{pq}|)^{(n+1)/2}} \tag{78}$$

where c_n^0 is the constant (40). However there is an exact analogue to formula (68). It is the formula

$$f(z) = \int_{U_n^0} \frac{f(\zeta)\,dv}{(\det|1 - z\zeta^*|)^{(n+1)/2}}, \tag{79}$$

where formally,

$$dv = c_n^0 \frac{d\zeta_1 \cdots d\zeta_k}{(\det|\zeta_{pq}|)^{(n+1)/2}}. \tag{80}$$

In order to justify the analogy, we consider the group of affine transformations

$$\zeta \to u'\zeta u, \tag{81}$$

where u is an arbitrary (non-symmetric) unitary transformation and u' is its transposed. These transformations are a *transitive* group of motions, and the existence of such a group leads to many properties similar to those of a group space proper. In particular there exists a group-invariant volume, and it is again unique, thus leading to formula (79). We will not insist on details of the proof. Obviously Theorem 6 has also a strict analogue in the symmetric case.

PART III

GENERAL GROUP INVARIANT DOMAINS.

11. An abstract theorem

Let G denote a compact group of affine transformation operating on a fixed set of variables $t = (t_1, \cdots, t_k)$. An element of G will be denoted by the letter σ, and the actual transformation by $t \to \sigma t$. The transposed transformation will be denoted by σ'. Thus, if (t, u) is the inner product $t_1u_1 + \cdots + t_ku_k$, then $(\sigma t, u) = (t, \sigma' u)$. We term G a *comprehensive* group, if any polynomial in t which is absolutely invariant for all transposed transformations σ' is of necessity a constant. For instance, the unitary group U_n is a comprehensive group. In fact, if an element of this group is denoted by ζ and if $P(t)$ is a polynomial in the set of variables $t = \{t_{pq}\}$, and if $P(t)$ is group invariant, then

$$f(\zeta; t) = P(\zeta t) - P(t) \tag{82}$$

is identically 0 for $\zeta \in U_n$. We now consider the function

$$f(z; t) = P(zt) - P(t) \tag{83}$$

for an arbitrary (non-unitary) matrix z, and fixed numerical values of t. Being a polynomial in z, we can apply to it Theorem 3, and since its boundary values on U_n are zero, it vanishes identically in z, that is $P(zt) = P(t)$. If we put $t = 1$ (unit matrix) we obtain

$$P(z) = P(1) = \text{constant},$$

thus proving that the unitary group is comprehensive. In the case of the unitary group the property just stated is part of the so-called "unitarian trick,"[8] and our concept of a comprehensive group is a generalization of this property, by explicit definition, to general affine groups. We will now state a theorem in which this concept has a vital part.

THEOREM 7. *If a distributive operation from (all) polynomials to polynomials is commutative with partial differentiation and invariant under the transformations of a comprehensive group, then the operation is a constant multiple of the identity.*

If we denote the operation by

$$F(x) = Lf(\xi) \tag{84}$$

then the assumptions are

---

[8] H. WEYL, *The classical groups*, Princeton (1943), 177, Lemma (7.1.A).

(85) $$L(af + bg) = aLf + bLg$$

(86) $$\frac{\partial}{\partial x_p} F(x) = L \frac{\partial f}{\partial \xi_p}; \quad p = 1, \cdots, k,$$

(87) $$F(\sigma x) = Lf(\sigma\xi); \quad \sigma \epsilon G$$

and the conclusion is

(88) $$F(x) = c_0 f(x),$$

where c_0 is a constant. We denote the polynomial $L(\xi_1^{p_1} \cdots \xi_k^{p_k})$ by

(89) $$F_{p_1 \cdots p_k}(x)$$

and we introduce the formal power series

(90) $$F(x; t) = \sum_p \frac{F_{p_1 \cdots p_k}(x) t_1^{p_1} \cdots t_k^{p_k}}{p_1! \cdots p_k!}.$$

Assumption (86) is equivalent with the system of relations

(91) $$\frac{\partial F}{\partial x_p} = t_p F; \quad p = 1, \cdots, k.$$

If we introduce the new expression $G(x; t) = e^{-(x,t)}F(x, t)$, then these relations are $\dfrac{\partial G}{\partial x_p} = 0$. Therefore $G(x; t)$ is constant in x, and thus we obtain for the system of polynomials (89) the generating function

(92) $$F(x; t) = e^{(x,t)} A(t_1, \cdots, t_k)$$

where $A(t)$ is a formal power series, as yet unrestricted. In order to apply the critical assumption (87) we introduce in addition to the series (90) the formal series

(93) $$e^{(\xi,t)} = \sum \frac{\xi_1^{p_1} \cdots \xi_k^{p_k} t_1^{p_1} \cdots t_k^{p_k}}{p_1! \cdots p_k!},$$

and after extending our operation to formal power series with polynomials as coefficients we obviously have

$$F(x; t) = Le^{(\xi,t)}.$$

By including assumption (87) we obtain the chain of equalities

$$e^{(\sigma x,t)}A(t) = F(\sigma x; t) = Le^{(\sigma\xi,t)} = Le^{(\xi,\sigma' t)} = F(x; \sigma' t) = e^{(x,\sigma' t)}A(\sigma' t),$$

and hence the relation

$$A(t) = A(\sigma' t).$$

Since the group is comprehensive, this relation can hold only if the power series $A(t)$ reduces to a constant c_0, that is $F(x, t) = c_0 e^{(x,t)}$. Therefore,

$$F_{p_1 \cdots p_k}(x) = c_0 x_1^{p_1} \cdots x_k^{p_k},$$

and this is the conclusion (88). We observe that there is nothing in our assumptions to prevent the constant c_0 from having the value zero.

A seemingly very special type of distributive operations having property (86) is given by the expression

$$F(x) = \sum_m f(x + \xi^m)\mu_m,$$

where $\{\xi^m\}$ are fixed points, and $\{\mu_m\}$ are fixed constants. However a transition from sums to Stieltjes integrals will lead from the special type to what is well-nigh the most general type. If C is any bounded Borel set and $\omega(A)$ is a completely additive set function of its Borel subsets, then the expression

$$F(x) = \int_C f(x + \xi)\, d\omega(\xi) \tag{94}$$

will obviously define a distributive operation having property (86). It will be easily seen that in order to obtain property (87) it is sufficient to assume that the set C and the set function $\omega(A)$ shall be group invariant; that is $\sigma C = C$, and $\omega(\sigma A) \equiv \omega(A)$.

An important type of Stieltjes integrand is generated by an external differential form. We assume that C is an orientable compact m-dimensional differentiable manifold C_m, $0 < m < k$, and we assume that a differential form

$$d\omega(\xi) = \sum K_{p_1 \cdots p_m}(\xi)\, d\xi_{p_1} \cdots d\xi_{p_m} \tag{95}$$

is defined in a neighborhood D of C_m. The set function $\omega(A)$ is then defined as $\int_A d\omega(\xi)$. The set function will be group invariant if the manifold C and the differential form are group invariant. This being so, if the group G is a comprehensive group, then the conclusion of Theorem 7, if written explicitly, is the formula

$$\int_{C_m} f(x + \xi) \sum K_{p_1 \cdots p_m}(\xi)\, d\xi_{p_1} \cdots d\xi_{p_m} = c_0 f(x). \tag{96}$$

We will also write it more succinctly in the form

$$\int_{C_m} f(x + \xi)\, d\omega(\xi) = c_0 f(x). \tag{97}$$

The formula is valid for the class of polynomials and in most cases, for no class of function which is substantially larger. It is easy to state a further condition under which the constant c_0 is $\neq 0$; the condition being that the group G shall be *transitive* on C_m. In fact, if G is transitive, the set function $\omega(A)$ must be a numerical multiple of a group invariant volume, and c_0 is this multiple, thus being different from zero.

12. Cauchy's formula

If, for a fixed point x, we replace the variable ξ in (97) by the variable $\xi - x$ then the manifold C_m goes over into a translated manifold which we will denote by $C_m(x)$, and formula (97) goes over into

$$\int_{C_m(x)} f(\xi)\, d_\xi \omega(\xi - x) = c_0 f(x). \tag{98}$$

The integrand in formula (98) has the shape we are aiming at, however the contour $C_m(x)$ is dependent on x; whereas in a "Cauchy formula" as we understand the concept the contour must be independent of x. We are thus naturally led to seeking conditions under which

$$\left(\int_{C_m(x)} - \int_{C_m}\right) f(\xi)\, d_\xi \omega(\xi - x) = 0. \tag{99}$$

By assumption, the differential form (95) is defined in some neighborhood D of C_m. Hence there exists a neighborhood N of the origin, such that for $x \in N$, $C_m(x)$ is also situated in D. It is easily seen that $C_m(x) - C_m$ is the oriented boundary of an $(m + 1)$-dimensional complex B in D. If by the theorem of Stokes, we transform the left side of (99) into an integral over B, then the relation (99) will hold provided

$$df \cdot d_\xi \omega(\xi - x) + f \cdot (d_\xi \omega(\xi - x))' = 0. \tag{100}$$

Now, if the latter relation were to hold for all polynomials $f(\xi_1, \cdots, \xi_k)$ in the (real) variables $\xi_1, \cdots, \xi_k$, then it would hold locally for general differentiable functions $f(\xi)$ and this would imply that the differential form (95) is identically 0. In other words there exists no Cauchy formula in real variables.

However the situation is different for complex variables and the reader will have no difficulties in proving the following theorem.

THEOREM 8. *If $\zeta_1, \cdots, \zeta_k$ are complex variables and $\bar\zeta_1, \cdots, \bar\zeta_k$ are conjugate complex variables; if $0 \leqq m < k$; if the tensor*

$$K_{p_1 \cdots p_m}(\zeta; \bar\zeta), \qquad p_1 < p_2 < \cdots p_m$$

is defined and differentiable in the neighborhood D of a compact orientable differentiable manifold C_{k+m} of dimension $k + m$; if the external derivative of the differential form

$$d\omega(\zeta; \bar\zeta) = d\zeta_1 \cdots d\zeta_k \sum_p K_{p_1 \cdots p_m}(\zeta; \bar\zeta)\, d\bar\zeta_{p_1} \cdots d\bar\zeta_{p_m} \tag{101}$$

is identically zero; and if the manifold and the differential form are invariant under the transformations

$$\{\zeta \to \sigma\zeta; \ \bar\zeta \to \bar\sigma\bar\zeta\}$$

of a comprehensive group G; then the relation

$$c_0 f(z) = \int_{C_{k+m}} f(\zeta)\, d_\zeta \omega(\zeta - z; \bar\zeta - \bar z)$$

holds for all analytic polynomials $f(\zeta) \equiv f(\zeta_1, \cdots, \zeta_k)$.

If the group G is transitive on C_{k+m}, then the constant c_0 is $\neq 0$.

The case $m = 0$ is the case of complex variables at its purest, whereas the case $m = k - 1$ is nearest to the case of Green's formula for real variables. The latter case has been treated extensively in a previous paper,[9] and will not be commented upon. In the case $m = 0$, the tensor reduces to a scalar $K(\zeta; \bar{\zeta})$ and the differential form (101) will be

$$K(\zeta; \bar{\zeta})\, d\zeta_1 \cdots d\zeta_k .$$

Its derivative is identically zero whenever $K(\zeta; \bar{\zeta})$ is independent of the conjugate quantities $\bar{\zeta}_1, \cdots, \bar{\zeta}_k$ thus being an analytic function $K(\zeta) \equiv K(\zeta_1, \cdots, \zeta_k)$. Thus we obtain formula

$$c_0 f(z) = \int_{C_k} f(\zeta) K(\zeta - z)\, d\zeta_1 \cdots d\zeta_k$$

provided C_k and $K(\zeta)\, d\zeta_1 \cdots d\zeta_k$ are invariant for some comprehensive group in the complex variables given. This formula obviously includes formulas (63) and (68).

Addendum

Another Special Case

We consider in the space of the real variables $u_1, \cdots, u_k$, $k \geqq 2$, the semi-cone

$$(P) \qquad u_1 > (u_2^2 + \cdots + u_k^2)^{1/2}.$$

The square-root is positive and therefore u_1 is automatically positive. The other variables are not algebraically restricted. It is easily seen that the relation

$$\alpha_1 u_1 + \cdots + \alpha_k u_k > 0$$

will hold for all u in P if and only if

$$(G) \qquad \alpha_1 > (\alpha_2^2 + \cdots + \alpha_k^2)^{1/2}.$$

Furthermore, by a known formula[10],

$$\int_G e^{-(\alpha_1 w_1 + \cdots + \alpha_k w_k)}\, dv_\alpha = \frac{2\pi^{(k+1)/2}\Gamma(k-1)}{\Gamma\left(\frac{k+1}{2}\right)\cdot \{w_1^2 - w_2^2 - \cdots - w_k^2\}^{k/2}}.$$

Thus we obtain the formula

$$F(w) = \frac{\Gamma\left(\frac{k}{2}\right)}{2\pi^{(k/2)+1} i^k} \int_{iE_k} \frac{F(\omega)\, d\omega_1 \cdots d\omega_k}{[(w_1 - \omega_1)^2 - (w_2 - \omega_2)^2 - \cdots - (w_k - \omega_k)^2]^{k/2}}$$

for functions of integrable square in the tube with basis P.

---

[9] S. Bochner, *Analytic and meromorphic continuation by means of Green's formula*, Annals of Math. 44 (1943), 652–673.

[10] S. Bochner, Vorlesungen ueber Fouriersche Integrale, p. 189.

Following E. Cartan[11] we now introduce new variables by the relations

$$w_1 = \frac{2z_1}{z_1^2 + \cdots + z_k^2} - 1$$

$$w_p = \frac{2iz_p}{z_1^2 + \cdots + z_k^2}, \qquad p = 2, \cdots, k,$$

and this leads to the formula

$$f(z) = \frac{\Gamma\left(\frac{k}{2}\right)}{2\pi i \, \pi^{k/2}} \int_U \frac{f(\zeta)\, d\zeta_1 \cdots d\zeta_k}{[(z_1 - \zeta_1)^2 + \cdots + (z_k - \zeta_k)^2]^{k/2}}.$$

The characteristic manifold U is described by

$$|\,\zeta_1^2 + \cdots + \zeta_k^2 - \zeta_1\,| = |\,\zeta_1\,|$$
$$\zeta_p(\bar{\zeta}_1^2 + \cdots + \bar{\zeta}_k^2) = \bar{\zeta}_p(\zeta_1^2 + \cdots + \zeta_k^2), \qquad p = 2, \cdots, k,$$

and the domain of validity in z is described by

$$|\,z_1^2 + \cdots + z_k^2 - z_1\,| < |\,z_1\,|$$

$$2(|\,z_1 - 1\,|^2 + |\,z_2\,|^2 + \cdots + |\,z_k\,|^2) \leqq 1 + |\,(z_1 - 1)^2 + z_2^2 + \cdots + z_k^2\,|^2.$$

This domain can be proven bounded.

PRINCETON UNIVERSITY

---

[11] Domaines borne's homogènes de l'espace de n variables complexes. Hamburgische Abhandlungen, 11 (1936), p. 149.

ANNALS OF MATHEMATICS
Vol. 45, No. 4, October, 1944

BOUNDARY VALUES OF ANALYTIC FUNCTIONS IN SEVERAL VARIABLES AND OF ALMOST PERIODIC FUNCTIONS

BY S. BOCHNER

(Received June 22, 1944)

INTRODUCTION

If an analytic function $f(z)$ in $|z| < 1$ is such that

$$\int_0^{2\pi} |f(re^{i\theta})| \, d\theta$$

is bounded for $0 < r < 1$ then there exists a boundary function $f(e^{i\theta})$ such that

$$\lim_{r\to 1} \int_0^{2\pi} |f(re^{i\theta}) - f(e^{i\theta})| \, d\theta = 0; \tag{1}$$

or if we "eliminate" the boundary function $f(e^{i\theta})$ then relation (1) is equivalent with the relation

$$\lim_{r\to 1, \rho\to 1} \int_0^{2\pi} |f(re^{i\theta}) - f(\rho e^{i\theta})| \, d\theta = 0.$$

This theorem of F. Riesz[1] has the distinction of being true only for analytic functions and not for harmonic functions in general as many similar theorems are. Correspondingly, the proof is very "analytic." It is based on the canonical decomposition

$$f(z) = g(z)\cdot B(z)$$

where $g(z) \neq 0$ and $B(z)$ has boundary values 1 almost everywhere (Blaschke product).

Now, for functions of several complex variables no comparable decomposition exists, and yet we will succeed in generalizing the theorem of Riesz proper. If $f(z_1, \cdots, z_k)$ is analytic for $0 \leqq |z_j| < 1, j = 1, \cdots, k$, and if

$$\int_0^{2\pi} \cdots \int_0^{2\pi} f(r_1 e^{i\theta_1}, \cdots, r_k e^{i\theta_k}) \, dv_\theta$$

is bounded then there exists an integrable boundary function $f(e^{i\theta_1}, \cdots, e^{i\theta_k})$ in

$$0 < \theta_j < 2\pi; \quad j = 1, \cdots, k \tag{2}$$

such that

$$\int_0^{2\pi} \cdots \int_0^{2\pi} |f(r_1 e^{i\theta_1}, \cdots, r_k e^{i\theta_k}) - f(e^{i\theta_1}, \cdots, e^{i\theta_k})| \, dv_\theta \tag{3}$$

[1] Ueber die Raudwerte einer analytischen Funktion, Mathem. Zeitschrift, 18 (1922), 87–95.

tends to 0 as $(r_1, \cdots, r_k) \to (1, \cdots, 1)$. The theorem can be stated entirely in terms of Fourier series. If we consider an additive set function of bounded variation on the Borel sets of the torus (2); if we introduce its Fourier series in the usual manner; and if the series has the special form

$$\sum_{m_1=0}^{\infty} \cdots \sum_{m_k=0}^{\infty} a_{m_1 \cdots m_k} e^{i(m_1\theta_1 + \cdots + m_k\theta_k)}$$

that is if the Fourier coefficient $a_{m_1 \cdots m_k}$ vanishes whenever at least one index m_j is negative; then the set function is of necessity absolutely continuous and therefore the indefinite integral of a Lebesgue point function. Actually we will show that for the conclusion to hold the non-vanishing Fourier coefficients need not be restricted rigidly to the "octant" $0 < m_j < \infty$ but may be contained in any convex cone which is strictly conical.

The theorem of Riesz can be deduced from a sharper theorem of Hardy-Littlewood[2] which states that the inequality

$$\int_0^{2\pi} |f(re^{i\theta})| \, d\theta < \gamma, \qquad 0 < r < 1$$

implies the sharper inequality

$$\int_0^{2\pi} \sup_{0<r<1} |f(re^{i\theta})| \, d\theta \leqq \alpha \cdot \gamma$$

where α is a universal constant. Now, if we replace the variable z by e^{-z} then the theorems of Riesz and Hardy-Littlewood are theorems on periodic analytic functions which are bounded in the half-plane $x > 0$. Our procedure will be to give at first a qualified extension of the theorem of Hardy-Littlewood to almost periodic functions of Bohr. This done, it will be quite easy to obtain the stated theorems in several variables by applying a fundamental relation which connects the two categories of functions.

After that we will turn our attention to almost periodic functions in their own right. We will give both an unequalified extension of the theorem of Hardy-Littlewood and a full generalization of the theorem of Riesz. The "boundary function" whose existence is involved in the theorem of Riesz will be a Besicovitch function. Nothing less general can be guaranteed. However the analytic function given in $x > 0$ may be a Besicovitch function to start with, and as a matter of fact we will set it up in a new scope of generality.

I. A Lemma on Mean Values

As a rule we will denote by $M_y f(y)$ the mean-value

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(y) \, dy$$

[2] A maximal theorem with function-theoretic applications. Acta Mathematica, 54 (1930), 81–116.

provided it exists. We will also write $\bar{M}_y(y)$ for the limit superior and $\underline{M}_y f(y)$ for the limit inferior, whenever $f(y)$ is real-valued.

THEOREM 1.[3] *If $K(y)$ is an even function in $(-\infty, \infty)$; if in $0 \leqq y < \infty$ $K(y)$ is differentiable and monotonely decreasing; and if*

$$\int_0^\infty K(y) < \infty,$$

thus implying that $yK(y) \to 0$ as $y \to \infty$; then for any non-negative function $\varphi(y)$ which is Lebesgue integrable over any finite interval we have the relations

$$\limsup_{\epsilon \to 0} \int_{-\infty}^{\infty} \varphi(y)\, \epsilon K(\epsilon y)\, dy \leqq \bar{M}_y \varphi(y) \cdot \int_{-\infty}^{\infty} K(y)\, dy \tag{4}$$

and

$$\underline{M}_y \varphi(y) \cdot \int_{-\infty}^{\infty} K(y)\, dy \leqq \liminf_{\epsilon \to 0} \int_{-\infty}^{\infty} \varphi(y)\, \epsilon K(\epsilon y)\, dy. \tag{5}$$

PROOF. Since $K(y)$ was assumed even, our relations may be proven in the form

$$\limsup_{\epsilon \to 0} \int_0^\infty \varphi(y)\, \epsilon K(\epsilon y)\, dy \leqq \bar{M}_y \varphi(y) \cdot \int_0^\infty K(y)\, dy \tag{6}$$

$$\underline{M}_y \varphi(y) \cdot \int_0^\infty K(y)\, dy \leqq \liminf_{\epsilon \to 0} \int_0^\infty \varphi(y)\, \epsilon K(\epsilon y)\, dy; \tag{7}$$

where $\varphi(y)$ is defined only in $0 \leqq y < \infty$ and where *for the remainder of this proof only*, the symbols $\bar{M}$ and $\underline{M}$ are the lim superior and lim inferior of $\frac{1}{T}\int_0^T \varphi(y)\, dy$ as $T \to \infty$.

The reader will next convince himself easily that we may now also assume that $\varphi(y)$ vanishes in a neighborhood of the origin. In particular, if we introduce the function

$$\Phi(y) = \frac{1}{y}\int_0^y \varphi(t)\, dt, \qquad 0 < y < \infty,$$

we may assume that

$$\Phi(y) \leqq \gamma, \qquad 0 < y < \infty.$$

Now

$$\int_0^\infty \varphi(y)\, \epsilon K(\epsilon y)\, dy = -\epsilon^2 \int_0^\infty yK'(\epsilon y)\Phi(y)\, dy$$

where $-K'(\epsilon y) \geqq 0$. Corresponding to any $\rho > 0$ there exists an A, such that

$$\Phi(y) \leqq \bar{M}_y \varphi(y) + \rho, \qquad A < y < \infty.$$

[3] Compare S. Bochner, Fouriersche Integrale, p. 30.

Therefore

$$-\epsilon^2 \int_0^\infty yK'(\epsilon y)\Phi(y)\, dy \leqq -\gamma \int_0^A \epsilon^2 yK'(\epsilon y)\, dy - (\bar{M} + \rho)\epsilon^2 \int_A^\infty yK'(\epsilon y)\, dy$$

$$\leqq -\gamma \int_0^{\epsilon A} yK'(y)\, dy + (\bar{M} + \rho) \int_0^\infty K(y)\, dy,$$

and this proves (6). On the other hand, for each $\rho > 0$ there exists a B such that

$$\Phi(y) \geqq \underline{M}_y\varphi(y) - \rho, \qquad B < y < \infty,$$

and (7) follows easily from

$$-\epsilon^2 \int_0^\infty yK'(\epsilon y)\Phi(y)\, dy \geqq -(\underline{M} - \rho)\epsilon^2 \int_B^\infty yK'(\epsilon y)\, dy.$$

II. A Theorem of Hardy-Littlewood

Theorem 2. *There exists an absolute constant α having the following properties.*
(i) *If*

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} | f(re^{i\theta}) | \, d\theta \leqq \gamma, \qquad 0 < r < 1,$$

then

(8) $$\frac{1}{2\pi} \int_{-\pi}^{\pi} \sup_{0<r<1} | f(re^{i\theta}) | \, d\theta \leqq \alpha \cdot \gamma.$$

(ii) *If*

(9) $$\int_{-\infty}^{\infty} | f(x + iy) | \, dy \leqq \gamma, \qquad 0 < x < \infty$$

then

(10) $$\int_{-\infty}^{\infty} \sup_{0<x<\infty} | f(x + iy) | \, dy \leqq \alpha \cdot \gamma.$$

(iii) *If*

(11) $$\frac{1}{2T} \int_{-T}^{T} | f(x + iy) | \, dy$$

is bounded in $0 < T < \infty$, $\xi < x < \infty$, for every fixed $\xi > 0$, and if

(12) $$\bar{M}_y \, | f(x + iy) | \leqq \gamma, \qquad 0 < x < \infty,$$

then for all $x_1 > x_0 > 0$ we have

(13) $$\underline{M}\{\max_{x_0 \leqq x \leqq x_1} | f(x + iy) |\} \leqq \alpha \cdot \gamma.$$

(iv) *The best constants α in each of the relations* (8), (10), (13) *have the same values.*

Proof. *Ad* (*i*). This is a renowned theorem of Hardy-Littlewood.[2] The

proof will not be repeated. The other parts of the theorem will be derived from (i).

Ad (*ii*). If $r = r(\theta)$ is a continuous function in $\theta_0 \leqq \theta \leqq \theta_1$, where $-\pi < \theta_0 < \theta_1 < \pi$, and if $0 < r(\theta) < 1$, then (8) implies

$$\frac{1}{2\pi}\int_{\theta_0}^{\theta_1} | f(r(\theta)e^{i\theta}) | \, d\theta \leqq \alpha \cdot \gamma. \tag{14}$$

Now, let $f(x + iy)$ be analytic in $x > 0$, and let

$$\{x = x(y);\ r \leqq y \leqq s\} \tag{15}$$

be a fixed rectilinear polygonal Jordan arc situated in a fixed rectangle

$$\{p \leqq x \leqq q;\ r \leqq y \leqq s\} \tag{16}$$

where

$$0 < p < q; \quad -\infty < r < s < \infty.$$

If $a > 0$ is sufficiently large, then the circle

$$| z - a | = a - \frac{p}{2}$$

will have the following properties. The curve (15) is contained in its interior, and any radius of the circle which meets the curve at some point meets it at one point only. Therefore, relation (14) implies the relation

$$\left(a - \frac{p}{2}\right)\int_r^s | f(x(y) + iy) | \frac{x'(y)\cdot y + a - x(y)}{(a - x(y))^2}\, dy$$
$$\leqq \alpha \int_{-\pi(a-p/2)}^{\pi(a-p/2)} \left| f\left(a + \left(a - \frac{p}{2}\right)e^{i\lambda/a-p/2}\right)\right| d\lambda.$$

We now let $a \to \infty$. The expression on the left converges to

$$\int_r^s | f(x(y) + iy) | \, dy.$$

In the integral on the right the path of integration converges to the straight line $x = \frac{p}{2}$, $-\infty < y < \infty$, and the integral will itself converge towards its formal limit which is

$$\int_{-\infty}^{\infty} \left| f\left(\frac{p}{2} + iy\right)\right| dy,$$

provided the function $f(z)$ in the half-plane $x > 0$ satisfies an estimate

$$| f(z) | \leqq \frac{M}{1 + | z |^3}, \qquad x > 0, \tag{17}$$

say. Thus for a function of this description the last inequality implies

$$\int_r^s |f(x(y) + iy)|\, dy \leqq \alpha \cdot \gamma. \tag{18}$$

Next, if $f(z)$ is an arbitrary function for which (9) holds then a sequence of functions $f_0(z)$ can be found which approximate to $f(z)$ in the norm (9) and such that each $f_0(z)$ satisfies an estimate (17) for some M. The sequence converges uniformly in (16), and thus (18) holds for $f(z)$ in general. We now hold the function $f(z)$ and the rectangle (16) fast, and vary the curve. By convergent sequences of curves we can reach any semi-continuous functions $y(x)$, and therefore (18) implies

$$\int_r^s \max_{p \leqq x \leqq q} |f(x + iy)|\, dy \leqq \alpha \cdot \gamma.$$

Letting $p \to 0$, $q \to +\infty$, $r \to -\infty$, $s \to +\infty$ we finally obtain (10).

Ad (*iii*). For $0 < \epsilon < \epsilon_0$ we introduce the function

$$g(z) = \frac{f(z)}{(1 + \epsilon z)^2}. \tag{19}$$

Since (19) is bounded for $x > \xi (> 0)$ we see that

$$\int_{-\infty}^{\infty} |g(x + iy)|\, dy$$

is bounded for $x > \xi$, and by (ii) we obtain for $x_1 > x_0 > \xi$ the inequality

$$\int_{-\infty}^{\infty} \max_{x_0 \leqq x \leqq x_1} |g(x + iy)|\, dy \leqq \alpha \int_{-\infty}^{\infty} |g(\xi + iy)|\, dy.$$

Therefore,

$$\int_{-\infty}^{\infty} \frac{\max_{x_0 \leqq x \leqq x_1} |f(x + iy)|}{(1 + \epsilon_0 x_0)^2 + \epsilon^2 y^2}\, dy \leqq \alpha \int \frac{|f(\xi + iy)|}{1 + \epsilon^2 y^2}\, dy.$$

We now multiply both sides by ϵ and apply Theorem 1. If we put $K(y) = (1 + y^2)^{-1}$ in (4) we obtain

$$\limsup_{\epsilon \to 0} \epsilon \int_{-\infty}^{\infty} \frac{|f(\xi + iy)|}{1 + \epsilon^2 y^2}\, dy \leqq \bar{M}_y |f(\xi + iy)| \cdot \pi,$$

and if we put $K(y) = [(1 + \epsilon_0 x_0)^2 + y^2]^{-1}$ in (5) we obtain

$$\underline{M}_y \max_{x_0 \leqq x \leqq x_1} |f(x + iy)| \frac{\pi}{1 + \epsilon_0 x_0}$$

$$\leqq \limsup_{\epsilon \to 0} \epsilon \int_{-\infty}^{\infty} \frac{\max_{x_0 \leqq x \leqq x_1} |f(x + iy)|\, dy}{(1 + \epsilon_0 x_0)^2 + \epsilon^2 y^2}.$$

If we put together the last three relations and let $\epsilon_0 \to 0$ we finally obtain (13).

Ad (iv). If α_1, α_2, α_3 are the best constants in (8), (10), (13) respectively, then our argument shows $\alpha_1 \geqq \alpha_2 \geqq \alpha_3$. However $\alpha_1 \leqq \alpha_3$ and thus $\alpha_1 = \alpha_2 = \alpha_3$, as asserted.

THEOREM 3. *If $f(z)$ is an analytic almost periodic function*[4] *in $x > 0$ [that is in each half plane $\xi \leqq x < \infty$, $\xi > 0$], and if*

$$M_y \mid f(x + iy) \mid \leqq \gamma, \qquad 0 < x < \infty, \tag{20}$$

then for each $x_0 > 0$ we have

$$M_y\{\sup_{x_0 \leqq x \leqq \infty} \mid f(x + iy) \mid\} \leqq \alpha \cdot \gamma. \tag{21}$$

PROOF. In fact, relation (13) is in the present case

$$M_y\{\max_{x_0 \leqq x \leqq x_1} \mid f(x + iy) \mid\} \leqq \alpha \cdot \gamma \tag{22}$$

for $x_1 > x_0 > 0$. Also, (20) implies that the exponents in the Dirichlet series of $f(z)$ are non-negative, and therefore for $x_1 \to \infty$ the integrand in (22) converges to the integrand in (21), uniformly in $-\infty < y < \infty$. This proves (21).

If $f(z)$ is pure-periodic we may pass from (21) to

$$M_y\{\sup_{0 < x < \infty} \mid f(x + iy) \mid\} \leqq \alpha \cdot \gamma \tag{23}$$

by ordinary Lebesgue theory. However, for almost periodic functions such an unqualified passage to the limit is not admissible without a detailed investigation, and the very existence of the mean-value (23) is a matter requiring investigation. We will not investigate this problem in the direct manner just outlined; but we will see in Part IV that there is a possibility of generalizing relation (23) to the almost periodic case.

III. FUNCTIONS IN SEVERAL VARIABLES

We will describe a well known connection between periodic functions in several variables and almost periodic functions in one variable. For fixed k, we consider the class of all continuous periodic functions

$$f(t_1, \cdots, t_k) \sim \sum a_{m_1 \cdots m_k} e^{i(m_1 t_1 + \cdots + m_k t_k)} \tag{24}$$

on the torus

$$0 \leqq t_j < 2\pi, \qquad j = 1, \cdots, k, \tag{25}$$

and we take a fixed set of positive real numbers

$$\mu_1, \cdots, \mu_k \tag{26}$$

which we assume to be linearly independent with regard to rational coefficients. Thus there is no relation of the form

$$g_1\mu_1 + \cdots + g_k\mu_k = 0$$

[4] As defined by H. Bohr, and not in any more general sense, for the present.

for integers $\{g_j\}$ other than $g_1 = \cdots = g_k = 0$. We now make the substitution

$$t_1 = \mu_1 t, \cdots, t_k = \mu_k t,$$

and the resulting function

$$g(t) = f(\mu_1 t, \cdots, \mu_k t) \tag{27}$$

is a Bohr almost periodic function with the expansion

$$\sum a_{m_1 \cdots m_k} e^{i(m_1\mu_1 + \cdots + m_k\mu_k)t}.$$

We will also denote the function $g(t)$ by

$$L(f) \equiv L(f; t).$$

The transformation

$$f \to Lf$$

is functional operation having the following properties. It is distributive, and it also transforms the real and imaginary parts of f separately into the real and imaginary parts of g. It is multiplicative,

$$L(f_1 \cdot f_2) = Lf_1 \cdot Lf_2,$$

and if $f_n \to f$ uniformly, then $Lf_n \to Lf$ uniformly. Therefore if

$$f_1(t), \cdots, f_n(t) \tag{28}$$

are real functions, and if

$$\lambda(u_1, \cdots, u_n)$$

is a continuous function in the real variables $u_1, \cdots, u_n$ which is defined and continuous on the closure of the trajectory (28), then

$$L(\lambda(f_1, \cdots, f_n)) = \lambda(Lf_1, \cdots, Lf_n). \tag{29}$$

Furthermore, the range of values of $g(t)$ is dense in the range of values of $f(t)$, and in particular,

$$\sup_t | g(t) | = \max_t | f(t_1, \cdots, t_k) |.$$

This will be of no consequence; but what will be of consequence is the relation

$$M_t g(t) = \frac{1}{(2\pi)^k} \int_0^{2\pi} \cdots \int_0^{2\pi} f(t_1, \cdots, t_k)\, dt_1 \cdots dt_k, \tag{30}$$

or written differently,

$$M_t L(f; t) = M_t f(t_1, \cdots, t_k).$$

Now, for $p \geqq 1$, and indeed for $p > 0$, the function

$$\lambda(u_1, u_2) = (u_1^2 + u_2^2)^{p/2}$$

is continuous. Therefore if we put $f = f_1 + if_2$ and apply (29) we see that

$$L(|f|^p) = |Lf|^p,$$

and if we substitute this in (31) we obtain

$$M_t |Lf|^p = M_t |f|^p. \tag{31}$$

Furthermore, the function

$$\lambda(u_1, \cdots, u_n) = \max (u_1, \cdots, u_n)$$

is continuous in $u_1 \geqq 0, \cdots, u_n \geqq 0$, therefore if functions

$$f_1, \cdots, f_n$$

are given, and if

$$g_\nu = Lf_\nu \qquad \nu = 1, \cdots, n$$

then

$$\max \{|g_1|, \cdots, |g_n|\} = L(\max \{|f_1|, \cdots, |f_n|\}),$$

and thus we obtain also

$$M_t \max \{|g_1|, \cdots, |g_n|\} = M_t \max \{|f_1|, \cdots, |f_n|\}. \tag{32}$$

We are now ready for the main theorem stated in the Introduction.

THEOREM 4. *If a power series*

$$f(s_1, \cdots, s_k) = \sum_{m_1=0}^{\infty} \cdots \sum_{m_k=0}^{\infty} a_{m_1 \cdots m_k} e^{-(m_1 s_1 + \cdots + m_k s_k)}$$

in the complex variables

$$s_j = \sigma_j + it_j, \qquad j = 1, \cdots, k,$$

is absolutely convergent in the domain

$$\sigma_1 > 0, \cdots, \sigma_k > 0 \tag{33}$$

and if for all $(\sigma_1, \cdots, \sigma_k)$ *in* (33) *we have*

$$\frac{1}{(2\pi)^k} \int_0^{2\pi} \cdots \int_0^{2\pi} |f(\sigma_1 + it_1, \cdots, \sigma_k + it_k)| \, dt_1 \cdots dt_k < \gamma \tag{34}$$

then the limiting series

$$\sum a_{m_1 \cdots m_k} e^{-i(m_1 t_1 + \cdots + m_k t_k)} \tag{35}$$

is the Fourier series of a function of class L^1.

More generally, if a Laurent series

$$f(s_1, \cdots, s_k) = \sum_{-\infty}^{\infty} \cdots \sum_{-\infty}^{\infty} a_{m_1 \cdots m_k} e^{-(m_1 s_1 + \cdots + m_k s_k)}$$

is absolutely convergent in a k-dimensional domain P of the $(\sigma_1, \cdots, \sigma_k)$*-space, if P is the union of half-lines issuing from but not containing the origin, and if relation* (34) *holds in P, then the limiting series is the Fourier series of a function in* L^1.

PROOF. Since P is a domain, and thus contains a neighborhood, it contains a point whose coordinates are linearly independent. Denote them by

$$\mu_1, \cdots, \mu_k.$$

By assumption, P contains the half-line

$$\sigma_j = \mu_j u, \qquad j = 1, \cdots, k; \qquad 0 < u < \infty.$$

Therefore we can introduce the function

$$g(w) = f(\mu_1 w, \cdots, \mu_k w), \qquad w = u + iv,$$

of the one complex variable w. This function is almost periodic, and, on the basis of formula (31), our relation (34) implies the kindred relation

$$M_v \mid g(u + iv) \mid \leqq \gamma, \qquad 0 < u < \infty, \tag{36}$$

since in terms of our previous operation we have

$$g(u + it) = Lf(\mu_1 u + it_1, \cdots, \mu_k u + it_k).$$

Now, take a sequence of numbers

$$u_1 > u_2 > u_3 > \cdots \to 0,$$

and consider the function

$$g_n(t) = g(u_n + it).$$

The latter is the transform of

$$f_n(t) = f(\mu_1 u_n + it_1, \cdots, \mu u_n + it_k).$$

By theorem 3, relation (36) implies

$$M_t \max \{ \mid g_1 \mid, \cdots, \mid g_n \mid \} \leqq \alpha \cdot \gamma,$$

and by relation (32) this is equivalent with

$$M_t \max \{ \mid f_1 \mid, \cdots, \mid f_n \mid \} \leqq \alpha \cdot \gamma, \tag{37}$$

$n = 1, 2, \cdots$. By ordinary Lebesgue theory on the torus (25) this implies that the sequence of functions $f_n(t_1, \cdots, t_k)$ is majorized by an integrable function on the torus (25). On the other hand, their Fourier series are convergent term-by-term to the limiting series (35). By a familiar lemma, the limiting series is the Fourier series of a function in L^1, and this completes the proof of theorem 4.

We note that in the first part of theorem 4, the assumption that our Laurent series shall be strictly a power series was redundant. It follows from the assump-

tion that relation (34) shall hold in the entire "octant" (33). Similar restrictions on the power series are implied in relation (34) depending on the type of domain P we choose in lieu of the octant (33). In fact, if P contains a half-line $(\sigma_1 u, \cdots, \sigma_k u)$, $0 < u < \infty$, then relation (34) implies that a coefficient $a_{m_1 \cdots m_k}$ can be different from zero only if

$$m_1\sigma_1 + \cdots + m_k\sigma_k \geqq 0. \tag{38}$$

This restriction is automatic. However if we are willing to introduce this restriction explicitly, then we may obtain another type of theorem.

THEOREM 5. *If (35) is the Fourier series of a set function of bounded variation on the Borel sets of the torus (25), and if the Fourier coefficients are $\neq 0$ only for such indices $m_1, \cdots, m_k$ as satisfy relation (38) for all $(\sigma_1, \cdots, \sigma_k)$ of some k-dimensional neighborhood N, no matter how small, then the set-function is absolutely continuous, and thus (35) is the Fourier series of a function in L^1.*

PROOF. We denote by P the point set consisting of all half-lines issuing from the origin and each containing at least one point of N. Obviously, (38) is satisfied for all points of P. We also introduce a sequence of exponential polynomials

$$f^{(\nu)}(t_1, \cdots, t_k) = \sum r^{(\nu)}_{m_1 \cdots m_k} a_{m_1 \cdots m_k} e^{-i(m_1 t_1 + \cdots + m_k t_k)}$$

which converge term by term towards (35) such that

$$M_t \,|\, f^{(\nu)}(t_1, \cdots, t_k) \,|\, \leqq \gamma,$$

where γ is the variation of our set function. From the restriction on the Fourier coefficients it follows that we also have

$$M_t \,|\, \sum r^{(\nu)}_m a_m e^{-(m_1 s_1 + \cdots + m_k s_k)} \,|\, \leqq \gamma \tag{39}$$

for $(\sigma_1, \cdots, \sigma_k)$ in P. We can now let $\nu \to \infty$ in (39), and then apply theorem 4.

It should be noted that theorem 5 is quite sharp; it is not permissible to let the neighborhood N shrink to a point. In fact, for $k = 2$, if

$$\sum a_{m_1, 0} e^{-i(m_1 t_1 + 0 \cdot t_2)}$$

is the Fourier series of a set function in one variable which is not absolutely continuous, then it is the Fourier series of a set function in two variables which is not absolutely continuous, although

$$\sigma_1 m_1 + \sigma_2 0 \geqq 0$$

for $\sigma_1 = 0$, $\sigma_2 = 1$ and all integers m_1.

REMARK ON L^p, $p > 1$. The latter comment emphasizes the distinction between our theorem concerning L^1, and a theorem of M. Riesz concerning L^p, $p > 1$. The theorem of M. Riesz states that if

$$\sum_{-\infty}^{\infty} a_m e^{-imt}$$

is the Fourier series of a function in L^p, then so is its "portion"

$$\sum_{m=0}^{\infty} a_m e^{-imt}.$$

As we have demonstrated somewhere else,[5] its generalization is that whenever (35) is a series in L^p, then so is its "portion"

$$\sum{}_{\sigma_1 m_1 + \cdots + \sigma_k m_k + \sigma_0 > 0},$$

this portion arising by dissection of the original series along any single one hyperplane. The fact is that the theorems of M. Riesz on conjugates is a theorem on functions in real variables (some of its proofs notwithstanding), whereas the theorem we are discussing in the present paper is not so.

IV. Almost Periodic Functions

We will consider in $-\infty < y < \infty$ almost periodic functions $f(y)$ of the most general kind, such being the functions of the Besicovitch class B^1. They result from closing the space of Bohr functions by the norm $M_y \mid f(y) \mid$. We will say that the functions $f(y)$, $g(y)$ are *equivalent* if

$$M_y \mid f(y) - g(y) \mid = 0.$$

In many important respects equivalent functions cannot be distinguished from one another. If $f(y)$ and $g(y)$ are both real, then we will say that $g(y)$ *exceeds* $f(y)$ if the difference $g(y) - f(y)$ is equivalent, in the sense just described, with a non-negative function. Investigations of almost periodic functions on general groups[6] have revealed many properties of Besicovitch functions analogous to those of periodic functions. In particular we will require the following two properties which are implied in general results.

Property 1. *If a sequence of real functions in B^1 are monotonely increasing, $g_n(y) \leqq g_{n+1}(y)$, and if their norms are bounded from above,*

$$M_y g_n(y) \leqq \gamma, \qquad n = 1, 2, \cdots,$$

then there exists a function $g(y)$ in B^1 which exceeds all $g_n(y)$ and for which

$$M_y g(y) \leqq \gamma.$$

Property 2. *If $\{f_n(y)\}$ is a sequence of complex functions in B^1, if all $\mid f_n(y) \mid$ are exceeded by some $g(y)$ in B^1, and if, for $n \to \infty$, the Fourier series of the sequence $f_n(y)$ are convergent term-by-term towards a limiting trigonometric series, then the latter series is again the Fourier series of some function in B^1.*

There is no clear-cut definition of functions B^1 of the complex variable $z =$

---

[5] S. Bochner, Additive set functions on groups, Annals of Math. 40 (1939), 769–799.

[6] In addition to the author's paper quoted in 5) see E. van Kampen, Almost periodic functions and compact groups, Annals of Math. 37 (1936), 78–91.

tion that relation (34) shall hold in the entire "octant" (33). Similar restrictions on the power series are implied in relation (34) depending on the type of domain P we choose in lieu of the octant (33). In fact, if P contains a half-line $(\sigma_1 u, \cdots, \sigma_k u)$, $0 < u < \infty$, then relation (34) implies that a coefficient $a_{m_1 \cdots m_k}$ can be different from zero only if

$$m_1\sigma_1 + \cdots + m_k\sigma_k \geqq 0. \tag{38}$$

This restriction is automatic. However if we are willing to introduce this restriction explicitly, then we may obtain another type of theorem.

THEOREM 5. *If (35) is the Fourier series of a set function of bounded variation on the Borel sets of the torus (25), and if the Fourier coefficients are $\neq 0$ only for such indices $m_1, \cdots, m_k$ as satisfy relation (38) for all $(\sigma_1, \cdots, \sigma_k)$ of some k-dimensional neighborhood N, no matter how small, then the set-function is absolutely continuous, and thus (35) is the Fourier series of a function in L^1.*

PROOF. We denote by P the point set consisting of all half-lines issuing from the origin and each containing at least one point of N. Obviously, (38) is satisfied for all points of P. We also introduce a sequence of exponential polynomials

$$f^{(\nu)}(t_1, \cdots, t_k) = \sum r^{(\nu)}_{m_1 \cdots m_k} a_{m_1 \cdots m_k} e^{-i(m_1 t_1 + \cdots + m_k t_k)}$$

which converge term by term towards (35) such that

$$M_t \,| f^{(\nu)}(t_1, \cdots, t_k) | \leqq \gamma,$$

where γ is the variation of our set function. From the restriction on the Fourier coefficients it follows that we also have

$$M_t | \sum r^{(\nu)}_m a_m e^{-(m_1 s_1 + \cdots + m_k s_k)} | \leqq \gamma \tag{39}$$

for $(\sigma_1, \cdots, \sigma_k)$ in P. We can now let $\nu \to \infty$ in (39), and then apply theorem 4.

It should be noted that theorem 5 is quite sharp; it is not permissible to let the neighborhood N shrink to a point. In fact, for $k = 2$, if

$$\sum a_{m_1, 0} e^{-i(m_1 t_1 + 0 \cdot t_2)}$$

is the Fourier series of a set function in one variable which is not absolutely continuous, then it is the Fourier series of a set function in two variables which is not absolutely continuous, although

$$\sigma_1 m_1 + \sigma_2 0 \geqq 0$$

for $\sigma_1 = 0$, $\sigma_2 = 1$ and all integers m_1.

REMARK ON L^p, $p > 1$. The latter comment emphasizes the distinction between our theorem concerning L^1, and a theorem of M. Riesz concerning L^p, $p > 1$. The theorem of M. Riesz states that if

$$\sum_{-\infty}^{\infty} a_m e^{-imt}$$

is the Fourier series of a function in L^p, then so is its "portion"

$$\sum_{m=0}^{\infty} a_m e^{-imt}.$$

As we have demonstrated somewhere else,[5] its generalization is that whenever (35) is a series in L^p, then so is its "portion"

$$\sum_{\sigma_1 m_1 + \cdots + \sigma_k m_k + \sigma_0 > 0},$$

this portion arising by dissection of the original series along any single one hyperplane. The fact is that the theorems of M. Riesz on conjugates is a theorem on functions in real variables (some of its proofs notwithstanding), whereas the theorem we are discussing in the present paper is not so.

IV. Almost Periodic Functions

We will consider in $-\infty < y < \infty$ almost periodic functions $f(y)$ of the most general kind, such being the functions of the Besicovitch class B^1. They result from closing the space of Bohr functions by the norm $M_y \,|\, f(y) \,|$. We will say that the functions $f(y)$, $g(y)$ are *equivalent* if

$$M_y \,|\, f(y) - g(y) \,| = 0.$$

In many important respects equivalent functions cannot be distinguished from one another. If $f(y)$ and $g(y)$ are both real, then we will say that $g(y)$ *exceeds* $f(y)$ if the difference $g(y) - f(y)$ is equivalent, in the sense just described, with a non-negative function. Investigations of almost periodic functions on general groups[6] have revealed many properties of Besicovitch functions analogous to those of periodic functions. In particular we will require the following two properties which are implied in general results.

Property 1. *If a sequence of real functions in B^1 are monotonely increasing, $g_n(y) \leqq g_{n+1}(y)$, and if their norms are bounded from above,*

$$M_y g_n(y) \leqq \gamma, \qquad n = 1, 2, \cdots,$$

then there exists a function $g(y)$ in B^1 which exceeds all $g_n(y)$ and for which

$$M_y g(y) \leqq \gamma.$$

Property 2. *If $\{f_n(y)\}$ is a sequence of complex functions in B^1, if all $|\, f_n(y) \,|$ are exceeded by some $g(y)$ in B^1, and if, for $n \to \infty$, the Fourier series of the sequence $f_n(y)$ are convergent term-by-term towards a limiting trigonometric series, then the latter series is again the Fourier series of some function in B^1.*

There is no clear-cut definition of functions B^1 of the complex variable $z =$

[5] S. Bochner, Additive set functions on groups, Annals of Math. 40 (1939), 769–799.

[6] In addition to the author's paper quoted in 5) see E. van Kampen, Almost periodic functions and compact groups, Annals of Math. 37 (1936), 78–91.

$x + iy$. All that can be required without prejudice is that the function shall have a Dirichlet expansion of the form

$$\sum a_n e^{-\lambda_n z}, \tag{40}$$

in the sense that for all x of some interval it shall be a B^1-function in the y-variable having the Fourier series

$$\sum_n (a_n e^{-\lambda_n x}) e^{-i\lambda_n y}. \tag{41}$$

Now, for our purpose it will be best to require just that and nothing more. Thus we start from a formal series of the type (40), and we assume that for each x in $0 < x < \infty$, the series (41) is the Fourier series of some element in B^1. Denoting this element by $f(x; y)$ we obtain a family of functions, where the variable x is a parameter designating membership in the family. We will call $f(x; y)$ a *formal* function.

THEOREM 6. *If the formal function $f(x; y)$ in $0 < x < \infty$ satisfies the inequality*

$$M_y \,|\, f(x; y) \,|\, \leqq \gamma, \qquad 0 < x < \infty, \tag{42}$$

then there exists a B^1-function $g(y)$ which exceeds $|\, f(x; y) \,|$ for every x, such that

$$M_y g(y) \leqq \alpha \cdot \gamma, \tag{43}$$

where α is the universal constant of theorem 2.

PROOF. We associate with (40) a sequence of approximating exponential polynomials

$$f^{(k)}(x; y) = \sum r_n^{(k)} e^{-\lambda_n z} \tag{44}$$

of the Fejer-type.[7] They have the properties

$$\lim_{k\to\infty} M_y \,|\, f^{(k)}(\xi; y) - f(\xi; y) \,|\, = 0, \qquad 0 < \xi < \infty \tag{45}$$

$$M_y \,|\, f^{(k)}(\xi; x) \,|\, \leqq \gamma, \qquad k = 1, 2, \cdots. \tag{46}$$

On account of (42) we may assume that all λ_n are non-negative, and therefore, by a known theorem[8] we have

$$M_y \,|\, f^{(k)}(\xi; y) \,|\, \geqq M_y \,|\, f^{(k)}(x; y) \,| \tag{47}$$

and

$$M_y \,|\, f^{(l)}(\xi; y) - f^{(k)}(\xi; y) \,|\, \geqq M_y \,|\, f^{(l)}(x; y) - f^{(k)}(x; y) \,| \tag{48}$$

for

$$\xi < x.$$

Letting $l \to \infty$, this implies

$$M_y \,|\, f(\xi; y) - f^{(k)}(\xi; y) \,|\, \geqq M_y \,|\, f(x; y) - f^{(k)}(x; y) \,| \tag{49}$$

[7] A. S. Besicovitch, Almost periodic functions, 1932, p. 103.

[8] Hardy, Ingham and Polya, Theorems concerning mean values of analytic functions, Proceedings of Royal Soc. Lond. 113 (1927), 542–569.

for

$$0 < \xi < x.$$

Since the function (44) is uniformly continuous in any strip we hence conclude the continuity property

$$\lim_{x\to\xi} M_y \,|\, f(\xi; y) - f(x; y) \,| = 0 \tag{50}$$

for $\xi > 0$.

We now apply theorem 3 in the half-plane $x > \xi$, instead of $x > 0$. In particular, if $x_1, \cdots, x_\rho$ are real numbers, all $> \xi$, then

$$M_y \max_{1\leq\sigma\leq\rho} |\, f^{(k)}(x_\sigma; y) \,| \leqq \alpha\cdot\gamma. \tag{51}$$

From the expression

$$\max (u_1, u_2) = \frac{u_1 + u_2 + |u_1 - u_2|}{2}$$

and from similar expressions for $\max (u_1, \cdots, u_\rho)$ the reader will conclude that we may let $k \to \infty$ in (51) and hence obtain

$$M_y g_\rho(y) \leqq \alpha\cdot\gamma, \qquad \rho = 1, 2, \cdots, \tag{52}$$

where

$$g_\rho(y) = \max_{1\leq\sigma\leq\rho} |\, f(x_\sigma; y) \,| \tag{53}$$

is a B^1-function each. We now take a sequence of numbers $\{x_\rho\}$ which is a dense set in $0 < x < \infty$, and we form the sequence of functions (53). They are increasing, and, by (52), their norms are bounded. By Property 1 of B^1-functions, there exists a limit-function $g(y)$ which exceeds all $g_\rho(y)$ and hence all $|\, f(x_\rho; y) \,|$. Thus all functions

$$g(y) - |\, f(x_\rho; y) \,|$$

exceed the function zero. By (50) it is now possible to conclude that $g(y) - |\, f(\xi; y) \,|$ exceeds zero for all $\xi > 0$, and this completes the proof of our theorem.

THEOREM 7. *If $f(x; y)$ is a formal function with expansion* (40) *and if* (42) *holds, then there exist a B^1-function $f(0; y)$ with expansion*

$$\sum a_n e^{-i\lambda_n y} \tag{54}$$

such that

$$\lim_{x\to 0} M_y \,|\, f(0; y) - f(x; y) \,| = 0. \tag{55}$$

PROOF. Since the functions $|\, f(x; y) \,|$ are exceeded by a function $g(y)$, the existence of a function with expansion (54) follows from Property 2 of B^1-functions. Relation (55) will then follow, if we approximate to $f(0; y)$ by our exponential polynomials

$$f^{(k)}(0; y) = \sum r_n^{(k)} a_n e^{-i\lambda_n y},$$

and then apply (48) and (49).

Remark on Functions B^p, $p > 1$. The following theorem is a secondary result since it applies to harmonic functions in general.

Theorem 8. *For $p > 1$, if $h(x; y)$ is a function of class B^p in $-\infty < y < \infty$ for $0 < x < \infty$, if its Fourier series has the form*

$$\sum e^{-\lambda_n x}(a_n \cos \lambda_n y + b_n \sin \lambda_n y)$$

and if

$$M_y \mid h(x; y) \mid^p \leqq \gamma, \qquad 0 < x < \infty,$$

then there exists a B^p-function $h(0; y)$ whose Fourier series is

$$\sum (a_n \cos \lambda_n y + b_n \sin \lambda_n y) \tag{56}$$

and

$$\lim_{x \to 0} M_y \mid h(0; y) - h(x; y) \mid^p = 0. \tag{57}$$

Proof. By a property of B^p-functions, the existence of the function $h(0; y)$ with series (56) follows directly from the facts that the family of functions $\{h(x; y)\}$ is bounded in norm and that their Fourier series are convergent term-by-term as $x \to 0$. Relation (57) follows again by applying algorithmic summation.

Finally we will demonstrate that there is no analogue to theorem 5 for almost periodic functions of unrestricted sequences of exponents $\{\lambda_n\}$.

A counter-example. We take a sequence of positive numbers $\mu_1, \mu_2, \mu_3, \cdots$ with the properties that no two general partial sums

$$\mu_{n_1} + \mu_{n_2} + \cdots + \mu_{n_r}$$

for $1 \leqq r < \infty$, $n_1 < n_2 < \cdots < n_r$, have the same numerical value. As for the existence of such sequences we note that any sequence of numbers which decrease to zero contains a subsequence of the given description. If we now form the functions

$$f_n(z) = \prod_{m=1}^{n} (1 + e^{-\mu_m z}),$$

then

$$M_y f_n(x + iy) = 1,$$

for all n and all x; also the expansions of $f_n(z)$ converge term-by-term, as $n \to \infty$, towards a series containing as sub-series the expansion

$$\sum_m e^{-\mu_m z}.$$

However, for $\mu_m \to 0$, the terms $\{e^{-\mu_m x}\}$ do not converge to zero for any x as $m \to \infty$, and thus the entire series cannot be the Fourier series of a B^1-function $f(x; y)$ for any real value x whatsoever.

Princeton University

Annals of Mathematics
Vol. 45, No. 4, October, 1944

TRANSFORMATION THEORY OF NON-LINEAR DIFFERENTIAL EQUATIONS OF THE SECOND ORDER

By Norman Levinson

(Received January 28, 1944)

1. Introduction

Interest in the field of non-linear differential equations in bygone years centered around the dynamics of conservative systems, the problems arising mainly in celestial mechanics. Current interest is focused mainly on non-conservative systems. Nevertheless many of the advances made in the past are of the greatest relevance to the problems of current interest. Therefore we shall give an account here of certain of these past results relevant to current problems. We shall also adapt certain methods so as to get results for these current problems.

So far as the non-linear equations that we shall consider are concerned, the most challenging problem mathematically appears to be that of finding means of excluding the possibility of certain singular situations. We shall describe some of these singular possibilities. As yet there is no method available to indicate under what conditions they cannot arise.

Transformation theory as a method in differential equations is due to Poincaré. The type of transformation we shall be interested in here is that of the Euclidean plane into itself. Certain methods and results of Birkhoff will be shown to be of interest.

In connection with their work on second order non-linear equations, the non-linear terms of which have a small parameter as a coefficient, Kryloff and Bogoliuboff[1] make use of transformation theory, particularly the theory that has been developed in the study of curves on a torus by Poincaré and Denjoy.

The second order equation

$$\ddot{x} + f(x, \dot{x})\dot{x} + g(x) = e(t), \tag{1.0}$$

with $e(t)$ of period L, is of considerable interest in applied mathematics. By setting $\dot{x} = y$ this equation becomes

$$\dot{x} = y, \qquad \dot{y} = -f(x, y)y - g(x) + e(t). \tag{1.1}$$

The system (1.1) is a particular case of

$$\dot{x} = F(x, y, t), \qquad \dot{y} = G(x, y, t), \tag{1.2}$$

where $F(x, y, t)$ and $G(x, y, t)$ are both periodic in t with period L. We shall assume that F and G are analytic in x, y, and t although much less would suffice for most of the discussion.

---

[1] N. M. Kryloff and N. N. Bogoliuboff, *Les Methodes de la Mécanique Non Linéaire Appliquées à la Théorie des Oscillations Stationnaires*, Monograph No. 8, Academy of Sciences of the Ukraines, 1934.

A system of the form (1.2) will be said to be of *class D, or to be a dissipative system for large displacements,* if there exists an R such that any solution $(x(t), y(t))$ of this system remains in the finite (x, y) plane as t increases and

$$\limsup_{t\to+\infty} x^2(t) + y^2(t) < R^2. \tag{1.3}$$

That is, any solution of a dissipative system of the form (1.2) eventually lies inside of a circle in the (x, y) plane with center at $(0, 0)$ and with radius depending only on the system. As might be expected from energy considerations, most systems which arise in practice are in class D.

This is a consequence of the following result: *Under rather general conditions*[2] *there exist simple closed curves in the (x, y) plane such that a solution $(x(t), y(t))$ of* (1.1) *can intersect any one of these curves only by crossing it from the domain exterior to the curve into the domain interior to the curve. Moreover through any point in the (x, y) plane sufficiently remote from the origin, there passes a curve with this property.*

Clearly if C_0 denote a simple closed curve of the type just described and if C_0 is sufficiently remote from the origin so that through every point exterior to it passes a curve of the type described above, then every solution of (1.1) starting in the domain exterior to C_0 must eventually cross into the interior of C_0 and stay there. Thus, under conditions generally met in practice, (1.1) is of class D.

It is well known that to a system such as (1.2) there corresponds a transformation of the (x, y) plane into itself. To see this let us consider the solution $(x(x_0, y_0, t), y(x_0, y_0, t))$ of (1.2) which when $t = t_0$ is at the point (x_0, y_0) of (x, y) plane. Let

$$x_n = x(y_0, t_0, t_0 + nL), \qquad y_n = y(x_0, y_0, t_0 + nL),$$

for any integer, n. Since F and G are of period L in t, it follows that

$$x_{n+m} = x(x_m, y_m, t_0 + nL), \qquad y_{n+m} = y(x_m, y_m, t_0 + nL). \tag{1.4}$$

If P_n denote the point (x_n, y_n), then we define a transformation T of the (x, y) plane into itself by $TP_0 = P_1$. By T^n is meant the transformation that takes P_0 into P_n. Clearly (1.4) is equivalent to $T^{n+m}P_0 = T^nP_m = T^nT^mP_0$. Moreover since solutions of (1.2) are continuous with respect to changes in initial values, T is continuous. If we now consider the solutions of (1.2) as curves in (x, y, t) space, it becomes clear that to study the behaviour of the solutions of (1.2) we have only to study the transformation T of the (x, y) plane into itself. In particular a fixed point under the transformation T^n corresponds to a periodic solution of (1.2) of period nL.

[2] See N. Levinson, *On the Existence of Periodic Solutions for Second Order Differential Equations with a Forcing Term.* Journal of Math. and Physics, Vol. XXII (1943), p. 41, Theorem II.

2. Maximum finite invariant domain

If we consider the closed curve C_0, described above, it follows that under iterations of T all points exterior to C_0 are transformed into interior points. Moreover if $T^n C_0$ be denoted by C_n, it follows that C_1 lies in the domain interior to C_0. Thus C_2 lies in the interior of C_1, etc. Consider the closed domain interior to all C_n, $n \geqq 0$. Clearly this domain is finite and is invariant under T. Moreover it is obviously the maximum finite invariant domain. We shall now prove the existence of such a maximum finite invariant domain for every system, (1.2) of class D.

THEOREM I. *Every system (1.2) of class D, possesses a closed domain, I, invariant under T. The complement of I is an open simply connected domain if the point at infinity is adjoined to the (x, y) plane. Under iterations of T every point in the complement of I except the point at infinity tends to I. I is the maximum finite invariant domain under T.*

THEOREM II. *Every system (1.2) of class D has at least one fixed point under the transformation T. Thus (1.2) has at least one solution of period L.*

PROOF OF THEOREM I. By the property, (1.3), of systems of class D and by the continuity of solutions of (1.2) with respect to changes in initial values, there exists an $N > 0$ such that if R_0 represents the circle, $x^2 + y^2 < R^2$, then the open simply connected domain $T^N R_0 = R_N$ is interior to R_0 and remains interior to R_0 in further applications of T. From $R_N \subset R_0$ it follows at once that $T^{2N} R_0 = R_{2N}$ is interior to R_N, etc. Let the closed domain interior to all R_{kN}, where $k = 0, 1, 2, \cdots$, be denoted by I.

Let Q_0 be an open simply connected domain in the interior of which is R_0 and R_1. Then by the property of systems of class D there must exist an n_0 such that $T^{n_0 N} Q_0 \subset R_0$. Thus we have $T^{n_0 N} Q_0 \subset R_0 \subset Q_0$. Applying T^{Nk} we have

$$R_{Nk} \subset Q_{Nk} \subset R_{(-n_0+k)N}. \tag{2.0}$$

Since $R_{Nk} \to I$ as $k \to \infty$, it follows that as $k \to \infty$

$$Q_{Nk} \to I \tag{2.1}$$

From $R_1 \subset Q_0$, it follows that $R_0 \subset Q_{-1}$. Proceeding with Q_{-1} as we have above with Q_0 we find that analogous to (2.0) we have

$$R_{Nk} \subset Q_{Nk-1} \subset R_{(-n_1+k)N}.$$

Thus $Q_{Nk-1} \to I$ as $k \to \infty$, or applying T we find $Q_{Nk} \to TI$. Combining this with (2.1) we have $TI = I$. Thus I is invariant under T. The other statements of Theorem I follow immediately and thus we have demonstrated the existence of the maximum finite invariant domain under T.

PROOF OF THEOREM II. As in the proof of Theorem I let the open domain, $x^2 + y^2 < R^2$, be denoted by R_0. Clearly $I \subset R_0$ and $R_{N+k} \subset R_0$, $k \geqq 0$. Consider the open set $R_0 + R_1 + \cdots + R_{N-1}$. A point P. is said to be occluded[3]

[3] G. D. BIRKHOFF. *Surface Transformations and Their Dynamical Applications.* Acta Mathematica, Vol. 43, 1922, see p. 80.

by this open set if a simple closed curve can be drawn entirely within this set and enclosing P and I. Let K_0 be the set of points occluded by this open set. Clearly K_0 is simply connected. Let $TK_0 = K_1$. Then K_1 is the set occluded by $R_1 + R_2 + \cdots + R_N$. Since $R_0 \supset R_N$, it follows that $K_1 \subset K_0$. Since R_0 is a circle and T analytic, it follows that the boundary of K_0 is everywhere accessible. (In fact it is made up of analytic curve segments.) Thus by the Brouwer fixed point theorem, there exists a point in K_0 fixed under T. Clearly this fixed point must be contained in I.

3. Invariant closed curves and rotation numbers

As with Birkhoff[4] we shall mean by a *closed curve* the boundary of a simply connected open continuum in the finite plane. The plane is regarded as completed by the adjunction of the point at infinity. A closed curve invariant under T is called an invariant closed curve. Clearly the boundary of the complement of I, the maximum finite invariant domain, is an invariant closed curve.

In addition there may be contained in I any number of invariant closed curves. In the simplest case where the right members of (1.2) actually do not involve t then any value of L is a period. In this case the invariant closed curves would be the limit cycles of Poincaré.

Let us denote an invariant closed curve by C. First let us consider the case where C is a simple closed curve of length λ. Consider all the solutions S_C of (1.2) starting on C when $t = t_0$. For $t_0 \leq t \leq t_0 + L$, these solutions, S_C, form a surface in (x, y, t) space which is bounded by C when $t = t_0$ and $t = t_0 + L$. Thus this surface can be mapped on the closed torus. The problem of the solutions, S_C, of (1.2) which emanate from C is reduced to the problem of the solution of a differential equation on a torus.

Clearly the equation of the surface formed by the solutions of (1.2) emanating from C can be written as $x = f(\theta, t)$ and $y = g(\theta, t)$ where f and g are of period 1 in θ and of period L in t. In fact f and g can be so chosen that, for any fixed $t = t_1$, θ is proportional to arc length on the curve, $x = f(\theta, t_1)$, $y = g(\theta, t_1)$. Differentiating x and y we find

$$\frac{dx}{dt} = \frac{\partial f}{\partial \theta}\frac{d\theta}{dt} + \frac{\partial f}{\partial t}, \qquad \frac{dy}{dt} = \frac{\partial g}{\partial \theta}\frac{d\theta}{dt} + \frac{\partial g}{\partial t}.$$

Using (1.2) and multiplying the first of the above equations by $\partial f/\partial\theta$ and the second by $\partial g/\partial\theta$ and adding we get

$$\frac{d\theta}{dt} = \frac{1}{\left(\frac{\partial f}{\partial \theta}\right)^2 + \left(\frac{\partial g}{\partial \theta}\right)^2}\left[F(f, g, t)\frac{\partial f}{\partial \theta} + G(f, g, t)\frac{\partial g}{\partial \theta} - \frac{\partial f}{\partial t}\frac{\partial f}{\partial \theta} - \frac{\partial g}{\partial t}\frac{\partial g}{\partial \theta}\right]. \tag{3.0}$$

Let λ_1 denote arc length on the curve $x = f(\theta, t_1)$, $y = g(\theta, t_1)$. Then since θ is proportional to arc length on each such closed curve we have

$$0 < \text{Min}\, \lambda_1^2 \leq \left(\frac{\partial f}{\partial \theta}\right)^2 + \left(\frac{\partial g}{\partial \theta}\right)^2 \leq \text{Max}\, \lambda_1^2 < \infty.$$

[4] Loc. cit. p. 79.

Thus the right member of (3.0) is bounded and is periodic of period 1 in θ and L in t. Its solutions as already stated are best represented geometrically on a torus.

Associated with such a differential equation, or with a family of curves on the torus, or with a transformation of a simple closed curve into itself is a rotation number, ρ.[5] This number is simply the average advance of θ, for an advance of $t = L$.

If ρ is rational and of the form p/q where p and q have no common factors, then (3.0) has solutions of period qL in t. Any non-periodic solution must tend toward such a periodic solution as $t \to \infty$. In this situation (1.2) has q^{th} subharmonics among its solutions.

If ρ is irrational there are two possibilities. One of these possibilities is termed the singular case. In case the right member of (3.0) is reasonably well behaved, for instance twice differentiable with respect to θ, the singular case is ruled out.[6] Under these latter circumstances the solutions of (3.0) are of the form

$$\theta(t) = \frac{t\rho}{L} + c + H\left(\frac{t}{L}, \frac{t\rho}{L} + c\right)$$

where[7] $H(u, v)$ is periodic in u and v of period 1 and c is an arbitrary constant. Clearly x and y are almost-periodic functions of t in this case. In fact they are of the form $h\left(\frac{t}{L}, \frac{t\rho}{L}\right)$ where $h(u, v)$ is of period 1 in u and v. In the case of certain systems which depart only slightly from linear Kryloff and Bogoliuboff[8] show that this is the case.

In the singular case there are solutions which trace out curves that together with their limit points intersect any meridian of the torus in a perfect, no-where-dense set of points.

There are cases of (1.2) of considerable practical interest where in effect, aside from a fixed point, there is only one invariant closed curve, C. If C were a reasonably well behaved curve the results just given would afford a complete qualitative solution of such cases. Even where there are several invariant closed curves the qualitative situation would be well in hand if these curves were known to be well behaved. Actually there is no indication that C need be a reasonable curve.

In fact Birkhoff[9] has given an example of an *analytic* transformation of the

---

[5] POINCARÉ. *Collected Works.* Vol. I, p. 137, Chapter XV.

[6] A. DENJOY. *Sur les courbes definies par les equations differentielles a la surface du tore.* Journal d. Math. Vol. 11 (1932), p. 333.

[7] P. BOHL. *Über die Hinsichtlich der Unabhängigen und Abhängigen Variabeln Periodische Differential Gleichungen. Erster Ordnung.* Acta Math. Vol. 40 (1916), p. 321.

[8] Loc. cit.

[9] G. BIRKHOFF. *Sur quelques Courbes Fermèes Remarquables.* Bull. de la Soc. Math. de France. Vol. 60, 1932, p. 1.

plane into itself which leaves a closed "curve," J, invariant. J divides the plane into two invariant open simply connected continua, S_i and S_e where to S_e is adjoined the point at infinity. Every point of J is a limit point of S_i or of S_e or of both. Those points of J accessible from S_e we denote by J_e, and from S_i by J_i. Associated with the transformation which carries J_e and J_i into themselves are rotation numbers ρ_e and ρ_i. What Birkhoff showed is that ρ_e need not equal ρ_i. This indicates how complicated J can be even though T is analytic. Unfortunately there is at present no basis for excluding curves of this type from the invariant closed curves of (1.2).

4. Fixed points

The solutions of (1.2) can be regarded geometrically as a family of curves in (x, y, t) space which, due to the periodicity of $F(x, y, t)$ and $G(x, y, t)$ in t need only be studied between the two planes $t = t_0$ and $t = t_0 + L$. Any point, (x_0, y_0), in the (x, y) plane when $t = t_0$ is carried into a point (x_1, y_1) in the (x, y) plane when $t = t_0 + L$ by the solution of (1.2) emanating from (x_0, y_0). This as we have seen defines the transformation T.

Let us consider the change in area in the (x, y) plane under the transformation T. Let $(x(x_0, y_0, t))$, $(y(x_0, y_0, t))$ denote the solution of (1.2) at (x_0, y_0) when $t = t_0$. Let the Jacobian of $(x(x_0, y_0, t), y(x_0, y_0, t))$ with respect to (x_0, y_0) be $\Delta(t)$. That is

$$\Delta(t) = \begin{vmatrix} \dfrac{\partial x}{\partial x_0} & \dfrac{\partial x}{\partial y_0} \\ \dfrac{\partial y}{\partial y_0} & \dfrac{\partial y}{\partial y_0} \end{vmatrix}. \tag{4.0}$$

Differentiating (4.0) and using (1.2) we find

$$\frac{d\Delta}{dt} = \left(\frac{\partial F}{\partial x} + \frac{\partial G}{\partial y}\right)\Delta.$$

Integrating and using $\Delta(t_0) = 1$, we find that $\Delta(t_0 + L)$, the Jacobian of the transformation T for the point (x_0, y_0) is given by

$$J\left(\frac{x_1, y_1}{x_0, y_0}\right) = \begin{vmatrix} \dfrac{\partial x_1}{\partial x_0} & \dfrac{\partial x_1}{\partial x_0} \\ \dfrac{\partial y_1}{\partial x_0} & \dfrac{\partial y_1}{\partial y_0} \end{vmatrix} = \exp\left[\int_{t_0}^{t_0+L}\left(\frac{\partial F}{\partial x} + \frac{\partial G}{\partial y}\right)dt\right]. \tag{4.1}$$

Thus the element of area $dx_0\,dy_0$ is carried by T into $J((x_1, y_1)/x_0, y_0))\,dx_0\,dy_0$.

Now let us study the transformation T in the neighborhood of a fixed point of the transformation. This will be facilitated by making a transformation of coordinates which takes the fixed point into the origin. That is let $(\bar{x}(t), \bar{y}(t))$ be a solution of (1.2) of period L. Then the point $(\bar{x}(t_0), \bar{y}(t_0))$ is a fixed point under T. Let P_0 be the point $(\bar{x}(t_0) + u_0, \bar{y}(t_0) + v_0)$ in the (x, y) plane. Denote TP_0 by P_1 with coordinates $(\bar{x}(t_0) + u_1, \bar{y}(t_0) + v_1)$. It is well known that the

solution $(x(t), y(t))$ starting at P_0 when $t = t_0$ can be represented by a power series in u_0 and v_0 with coefficients functions of t. Thus

$$x(t) = \bar{x}(t) + c_1(t)u_0 + c_2(t)v_0 + c_3(t)u_0^2 + c_4(t)u_0v_0 + \cdots$$

and similarly for $y(t)$. In particular setting $t = t_0 + L$ we get

$$u_1 = au_0 + bv_0 + \cdots \tag{4.2}$$
$$v_1 = cu_0 + dv_0 + \cdots$$

where the terms not explicitly given in the right members are of degree two or higher in u_0 and v_0.

If we denote $(\bar{x}(t_0) + u_0, \bar{y}(t_0) + v_0)$ by (x_0, y_0) and $(\bar{x}(t_0) + u_1, \bar{y}(t_0) + v_1)$ by (x_1, y_1) then clearly

$$J\left(\frac{x_1, y_1}{x_0, y_0}\right) = J\left(\frac{u_1, v_1}{u_0, v_0}\right).$$

Using (4.1) and (4.2) we have

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \exp\left[\int_{t_0}^{t_0+L}\left(\frac{\partial F(\bar{x}, \bar{y}, t)}{\partial x} + \frac{\partial G(\bar{x}, \bar{y}, t)}{\partial y}\right)dt\right]. \tag{4.3}$$

Thus $ad - bc > 0$.

The transformation (4.2) has been very much studied. For small values of u_0 and v_0, its character is determined by its linear terms. That is, the transformation can be characterized by the roots of the equation

$$(a - \rho)(d - \rho) - bc = 0. \tag{4.4}$$

If one or both roots of (4.4) is 1, this means that $(\bar{x}(t_0), \bar{y}(t_0))$ is a multiple fixed point of T. A slight change in the parameters of the equation (1.2) will ordinarily separate these points and thus the situation requires no special treatment but can be studied as a limiting case of simple fixed points. We shall therefore assume the roots of (4.4) to be different from 1 in the discussion that follows. We shall also assume the roots different from -1, which indicates multiple fixed points under T^2.

Let us denote the roots of (4.4) by ρ_1 and ρ_2. Since $ad - bc > 0$, it follows from (4.4) that $\rho_1\rho_2 > 0$. We now characterize the periodic solution, $(\bar{x}(t), \bar{y}(t))$, or what is the same, the fixed point, $(\bar{x}(t_0), \bar{y}(t_0))$, as *completely stable* if $|\rho_1| < 1$, $|\rho_2| < 1$. In this case the point (u_1, v_1) is nearer to $(0, 0)$ than (u_0, v_0). Iterations of T bring (u_0, v_0) nearer and nearer to $(0, 0)$. That is, all solutions of (1.2) near $(\bar{x}(t), \bar{y}(t))$ move closer and closer to $(\bar{x}(t), \bar{y}(t))$ at $t \to \infty$.

Similarly if $|\rho_1| > 1$, $|\rho_2| > 1$, then we shall term the solution $(\bar{x}(t), \bar{y}(t))$ *completely unstable.*

If $\rho_1 > 1 > \rho_2 > 0$, then we shall call the solution $(\bar{x}(t), \bar{y}(t))$, *directly unstable.*

If $\rho_2 < -1 < \rho_1 < 0$, then the solution $(\bar{x}(t), \bar{y}(t))$, will be called *inversely unstable.*

The only other possibility is the case $|\rho_1| = |\rho_2| = 1$. In this case $\rho_2 = \bar{\rho}_1$. This is a very important case in conservative systems. In this case stability can be investigated only by considering higher powers of u_0 and v_0 in (4.2). Since in this case stability is not determined by the linear part of the transformation we shall call it the *undetermined* case.

In the case of directly or inversely unstable points there is a curve invariant under T passing through the fixed point[10]. Points on this invariant curve move farther and farther away from the fixed point under iterations of T. There is another invariant curve passing through the fixed point, points of which move toward the fixed point under iterations of T. Aside from the case where ρ_1 or ρ_2 are ± 1, the above classification of stability is exhaustive.

5. The Fixed Point Equation

A fruitful tool in the investigation of the transformation, T, involves the use of the vector field in the (x, y) plane where from each point, P_0, there emanates a vector directed toward and terminating in $TP_0 = P_1$. The application we shall make here is a special case of results of Birkhoff.[11]

Clearly the number of revolutions made by P_0P_1 as P_0 traces out a closed curve in the (x, y) plane must be an integer since P_0 returns to its starting position. This integer is called the index of the curve. The curve, we assume does not pass through any fixed points of T.

We shall first show that the closed curve which is the boundary of K_0, the continuum introduced in the proof of Theorem II, has index 1. Clearly this curve is free of fixed points. Since $TK_0 \subset K_0$, the index would certainly be 1 if K_0 were the interior of a circle. For in this case the vector $\overrightarrow{P_0P_1}$ must make an angle of less than 90° with the radius from the center of the circle to P_0, and the radius makes one revolution when P_0 does.

We now consider the general case of K_0 and gradually deform it into the interior of a circle. It follows that since the index of the boundary changes continuously with the deformation and since it is an integer, it must be 1.

(We assume in what follows that there is no curve in the (x, y) plane each point of which is invariant under T.) We enclose each fixed point by a small circle. We join the circles together by curves so that if the curves be regarded as cuts, the curves and circles together form a single closed curve. Shrink the boundary of K_0 down to this closed curve. Since the index will change continuously, it must remain 1 for the single closed curve made up of the circles and several curves joining them. In determining the index of the single closed curve, each curve segment, or cut, joining a pair of circles is traversed first in one direction and then the other. Thus the net effect of these segments on the index is zero.

[10] J. Hadamard, *Sur l'iteration et les solutions asymptotiques des equations différentielles.* Bull. de la Soc. math. de France, Vol. 26, 1901. Actually there is an invariant curve if ρ_1 and ρ_2 are real and unequal.

[11] *Dynamical systems with two degrees of freedom.* Trans. of the Am. Math. Soc., vol. 18 (1917), p. 287. The method is due to Poincaré, Collected Works, Vol. 1, p. 25.

Therefore the sum of the indices of the several circles enclosing the fixed points of T must be 1.

Let P_0 be a point on the circumference of such a small circle. Let ϕ be the angle P_0P_1 makes with the x axis, or what is the same thing, the u axis. Then

$$\phi = \tan^{-1}\frac{y_1 - y_0}{x_1 - x_0} = \tan^{-1}\frac{v_1 - v_0}{v_1 - v_0}$$

or using (4.2)

$$\phi = \tan^{-1}\frac{cu_0 + (d-1)v_0 + \cdots}{(a-1)u_0 + bv_0 + \cdots}. \tag{5.0}$$

If the radius of the small circle is r and if the angle the radius to P_0 makes with the u axis is θ, then using (5.0) we get

$$\phi = \tan^{-1}\frac{c\cos\theta + (d-1)\sin\theta + r[\cdots]}{(a-1)\cos\theta + b\sin\theta + r[\cdots]}. \tag{5.1}$$

From (5.1)

$$\frac{d\phi}{d\theta} = \frac{(a-1)(d-1) - bc + r[\cdots]}{[(a-1)\cos\theta + b\sin\theta + \cdots]^2 + [\cos\theta + (d-1)\sin\theta + \cdots]^2}. \tag{5.2}$$

Since we have excluded the case where (4.4) has roots equal to 1,

$$(a-1)(d-1) - bc \neq 0.$$

Thus it is clear from (5.2) that $d\phi/d\theta$ is of fixed sign which it takes from $(a-1)(d-1) - bc$. Also the denominator of the right member of (5.1) vanishes twice as θ goes from 0 to 2π. Since ϕ is either monotonically increasing or decreasing, this means that ϕ changes by 2π or -2π according to the sign of $(a-1)(d-1) - bc$ when P_0 traverses the circle. In other words each circle enclosing a fixed point is of index equal to $[(a-1)(d-1) - bc]$. A consideration of the relationship of this sign to the location of the roots of $(a-\rho)(d-\rho) - bc$ reveals immediately that in the cases completely stable, completely unstable, inversely unstable, and undetermined, $(a-1)(d-1) - bc > 0$. In the case directly unstable, $(a-1)(d-1) - bc < 0$. Since the sum of the indices is 1, we have

THEOREM III. *If C denote the number of completely stable, completely unstable, and undetermined points of T, if I denote the number of inversely unstable points, and D the number of directly unstable points we have*

$$C + I = 1 + D. \tag{5.3}$$

Although (5.3) was proved for T, the proof would obviously apply to the transformation T^n where n is any integer. Thus (5.3) applies to the solutions of (1.2) of period nL.

6. A Theorem on Subharmonics

A solution of (1.2) is called subharmonic if its least period is equal to qL where $q > 1$ is an integer. We shall now show, for any q, that not all the subharmonic solutions of least period of qL can be stable. To be more specific we shall show

THEOREM IV. *If* (1.2) *is of class D and possesses N subharmonic solutions of least period qL, then $N = 2kq$, where k is an integer. Moreover kq, that is one half, of the solutions are directly unstable.*

Let $A(m)$ represent the number of completely stable, completely unstable, inversely unstable, and undetermined solutions of (1.2) of *least* period mL, and let $D(m)$ represent the number of directly unstable solutions of *least* period mL.

Clearly if m_1 is a factor of m, then the solutions of (1.2) of period m_1L appear among those of period mL. In other words among the fixed points of T^m are those of T^{m_1}. Let

$$m = p_1^{n_1}p_2^{n_2} \cdots p_n^{n_n}$$

where p_j, $j = 1, 2, \cdots, n$, are prime numbers. Then the total number of solutions of (1.2) of period mL is given by

$$\sum_{j_1=0}^{n_1}\sum_{j_2=0}^{n_2}\cdots\sum_{j_n=0}^{n_n}[A(p_1^{j_1}p_2^{j_2}\cdots p_n^{j_n}) + D(p_1^{j_1}p_2^{j_2}\cdots p_n^{j_n})].$$

Applying (5.3) to the transformation T^m we find

$$\sum_{j_1=0}^{n_1}\sum_{j_2=0}^{n_2}\cdots\sum_{j_n=0}^{n_n}[A(p_1^{j_1}p_2^{j_2}\cdots p_n^{j_n}) - D(p_1^{j_1}p_2^{j_2}\cdots p_n^{j_n})] = 1. \tag{6.0}$$

If we now apply (6.0) to the case where $m = p_1^{n_1-1}p_2^{n_2}\cdots p_n^{n_n}$ and subtract the equation so obtained from (6.0), we get

$$\sum_{j_2=0}^{n_2}\cdots\sum_{j_n=0}^{n_n}[A(p_1^{n_1}p_2^{j_2}\cdots p_n^{j_n}) - D(p_1^{n_1}p_2^{j_2}\cdots p_n^{j_n})] = 0. \tag{6.1}$$

If we consider (6.1) for $m = p_1^{n_1}p_2^{n_2-1}p_3^{n_3}\cdots p_n^{n_n}$ and subtract the result so obtained from (6.1) we get

$$\sum_{j_3=0}^{n_3}\cdots\sum_{j_n=0}^{n_n}[A(p_1^{n_1}p_2^{n_2}p_3^{j_3}\cdots p_n^{j_n}) - D(p_1^{n_1}p_2^{n_2}p_3^{j_3}\cdots p_n^{j_n})] = 0. \tag{6.2}$$

Proceeding in this way we finally get

$$A(p_1^{n_1}p_2^{n_2}\cdots p_n^{n_n}) = D(p_1^{n_1}p_2^{n_2}\cdots p_n^{n_n}).$$

In other words

$$A(m) = D(m). \tag{6.3}$$

Let $D(q)$ be the number of directly unstable fixed points under T^q but not fixed for T^n where $n > 0$ is less than q. Consider such a fixed point P. Then P, TP, T^2P, $\cdots T^{q-1}P$, are all distinct. Thus the fixed points under consideration

fall into mutually exclusive sets of q points. That is $D(q) = kq$ where k is an integer. Now from (6.3) it follows that $A(q) = D(q)$. Thus the total number of points fixed under T^q and not T^n, $0 < n < q$, is $D(q) + A(q) = 2kq$. This completes the proof of the theorem.

7. Some Examples

The simplest possible maximum finite invariant domain is that consisting of a single point. This is actually the case for the following two equations[12], (7.0) and (7.1):

$$\ddot{x} + f(x)\dot{x} + x = e(t) \tag{7.0}$$

where $e(t)$ is periodic, $xf(x) > 0$ for $x \neq 0$, and $\int^{\infty} f(x)\, dx = \infty$;

$$\ddot{x} + F(\dot{x}) + x = e(t) \tag{7.1}$$

where $F(y) > 0$ for $y \neq 0$ and $F(y) \to \infty$ as $y \to \infty$.

The differential equation

$$\ddot{x} + f(x, \dot{x})\dot{x} + g(x) = e(t),$$

under the conditions enumerated in the paper referred to in connection with this equation in §1, and with the further condition

$$f(x, y) + \frac{\partial f}{\partial y} y > 0, \qquad (y \neq 0, \quad x \neq 0), \tag{7.2}$$

has its maximum finite invariant domain of zero area. To see this we note that the left member of (7.2) is $-(\partial F/\partial x + \partial G/\partial y)$ in our general notation. Thus by (4.1) the Jacobian of T under these conditions is always less than 1. Using the notation in the proof of Theorem II we note that I is the inner limiting set of the domains enclosed by the curves K_n. Denoting by A_0 the area bounded by K_0,

$$\text{area of } I = \lim_{n\to\infty} \iint_{A_0} dx_0\, dy_0 \exp\left[-\int_{t_0}^{t_0+nL}\left(f(x, \dot{x}) + \dot{x}\frac{\partial f}{\partial x}\right) dt\right]. \tag{7.3}$$

But by (7.2), for all (x_0, y_0) in K_0 there exists a $\delta > 0$ such that

$$\int_{t_0}^{t_0+L}\left(f + x\frac{\partial f}{\partial x}\right) dt \geqq \delta.$$

Thus (7.3) becomes

$$\text{area of } I \leqq \lim_{n\to\infty} \iint_{A_0} dx_0\, dy_0\, e^{-n\delta} = 0.$$

---

[12] N. LEVINSON. *On a Non-Linear Differential Equation of the Second Order.* Journal of Math. and Physics, Vol. 22 (1943), p. 181.

The statement that I is of zero area clearly implies that every point of I is a limit point of the open continuum exterior to I. This case includes the important case of

$$\ddot{x} + \epsilon\dot{x} + g(x) = e(t), \qquad \epsilon > 0.$$

Being of zero areas, I is of a comparatively simple character and a number of special results can be given. For the present we point out a few simple possible types of I of zero area. The case where I is a single point, which arises for instance with equations (7.0) or (7.1), signifies that there exists one periodic solution of period L toward which all other solutions tend.

The case where I is a segment of a curve with three fixed points is of considerable interest. Let I consist of the curve segment ABC where A and C are the end points of the segment and A, B, and C are all fixed under T. A and C are completely stable fixed points, and B is directly unstable. ABC is shown in Fig. 1.

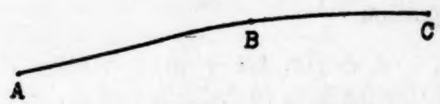


FIG. 1

In case parameters in the differential equation are varied, ABC will change its shape, but we shall suppose ABC remains topologically equivalent to a line segment. This turns out to be the case in many situations of practical importance. What does happen though, very often, is that as the parameters in the differential equation change B will move up gradually to the point C, and will actually come into coincidence with C. Then as the parameters are varied further, I will consist merely of the point A.

If we consider the periodic solution corresponding to the fixed point C, what we find is that this solution remains stable until B comes into coincidence with C, whereupon the only stable solution is the one corresponding to A. This situation is well known in practice as the "jump" phenomenon whereby a nonlinear system may with a small change in parameter jump from one steady state to another.

Another I of interest is shown in Fig. 2. Here 0 is the only fixed point under T and is completely stable. We have $TA = B$, $TB = C$, $TC = A$ and similarly $T^3D = T^2E = TF = D$. That is A, B, C, D, E, and F are fixed under T^3. Moreover under T^3, A, B, and C are completely stable and D, E, and F directly unstable. In this case the rotation number of I is $2\pi/3$. By Theorem IV we see that this case corresponds to the simplest I for which we can have a subharmonic of period $3L$.

Another I of interest is shown in Fig. 3. Here the portion of I from O through D continues on by wrapping itself an infinite number of times around the curve segment containing A. The same is true of the two other branches of I going

through E and F. As before $TA = B$, $TB = C$, $TC = A$ and similarly for D, E, and F. We also have $TG = H$, $TH = I$, $TI = J$, $TJ = K$, $TK = L$, and $TL = G$. That is, these points are fixed under T^6. Thus in this case A, B, and C are inversely unstable under T^3 and D, E, and F are directly unstable under T^3. Here although the rotation number is $2\pi/3$, we have no stable third

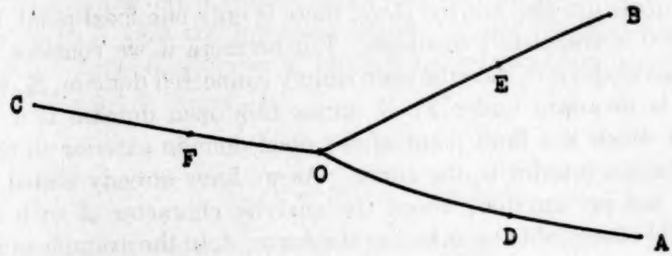


FIG. 2

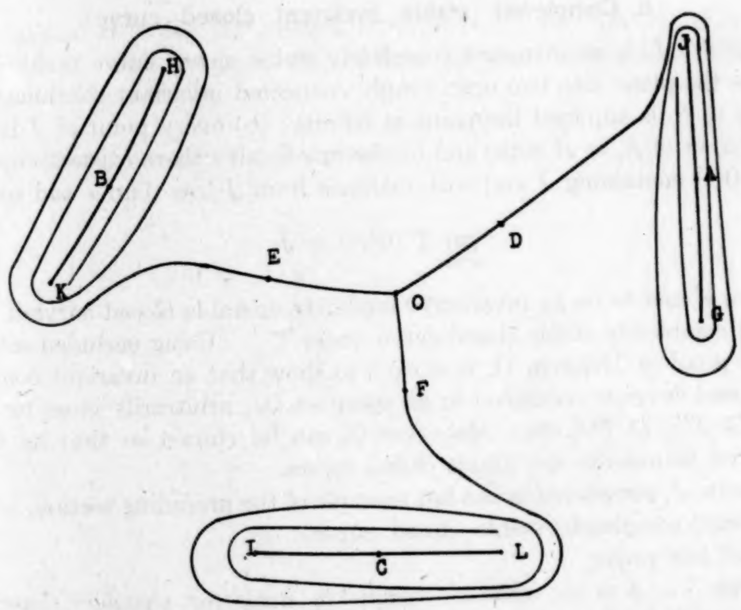


FIG. 3

order subharmonics. However corresponding to G we will have a stable subharmonic of order 6.

Another case, a singular case which we cannot rule out as a possibility, has an I which has only one fixed point, 0, under T or any iterate of T. Topologically I is equivalent to certain radii of a circle with 0 as its center. These radii emanate from 0 and terminate in a perfect no-where dense set of points on the circumference of this circle. A solution corresponding to a point of this perfect

set is of the type Birkhoff calls discontinuous recurrent. It would be very desirable to have a criterion for ruling out this type of I.

As a final example let us take the van der Pol equation with a forcing term

$$\ddot{x} + \mu(x^2 - 1)\dot{x} + x = e(t),$$

with μ large. In this case I is not of zero area. If $e(t)$ is not too large, and if the period of $e(t)$ is likewise not too large, there is only one fixed point, 0, under T. Moreover 0 is completely unstable. Furthermore if we consider a circle, S_0, of small radius about 0, then the open simply connected domain, $S_0 + TS_0 + T^2S_0 + \cdots$, is invariant under T. I minus this open domain is a "curve" every point of which is a limit point of the open domain exterior to the curve or the open domain interiot to the curve. As we have already stated, we can unfortunately not say anything about the analytic character of such a curve. This curve could conceivably be as bad as the curve, J, in the example of Birkhoff already cited.

8. Completely stable invariant closed curve

A "curve," J, is an invariant completely stable closed curve under T if (a) it divides the plane into two open simply connected invariant continua, S_i and S_e where to S_e is adjoined the point at infinity; (b) every point of J is a limit point of S_i or of S_e or of both; and (c) for any small ϵ there exists an open continuum, $0(\epsilon)$ containing J and with distance from J less than ϵ and such that

$$\lim_{n\to\infty} T^n(0(\epsilon)) = J.$$

A curve is said to be an invariant completely unstable closed curve if it is an invariant completely stable closed curve under T^{-1}. Using occluded sets much as in the proof of Theorem II, it is easy to show that an invariant completely stable closed curve is contained in an open set O_0, arbitrarily close to J, such that $O_0 \supset TO_0 \supset TO_0^2$ etc. Moreover O_0 can be chosen so that its exterior and interior boundaries are simple closed curves.

The curve, J, considered in the last example of the preceding section, is clearly an invariant completely stable closed curve.

We shall now prove

THEOREM V. *J is an invariant completely stable (or unstable) closed curve. If J contains N fixed points under T^q, $q \geqq 1$, but not fixed under T^j, $0 < j < q$, then $N = 2kq$ where k is an integer. Moreover kq of the fixed points are directly unstable.*

First consider the open simply connected domain consisting of O_0, described above, and all points in S_i, the interior of J. Let $A_0(q)$ be the number of completely stable, completely unstable, inversely unstable and undetermined points invariant under T^q, and no smaller power of T, and contained in the open simply connected domain under consideration. Similarly $D_0(q)$ is the number of di-

rectly unstable points. Then using exactly the same procedure as in the proofs of Theorem III and IV, we have

$$(8.0)\qquad \begin{aligned} A_0(1) &= 1 + D_0(1) \\ A_0(q) &= D_0(q), \qquad q > 1. \end{aligned}$$

Consider next the open simply connected domain bounded by the interior boundary of O_0. Call the invariant points for this domain $A_1(q)$ and $D_1(q)$, all these points being fixed under T^q but not under any smaller power of T. Now applying T^{-q} we get as before

$$(8.1)\qquad \begin{aligned} A_1(1) &= 1 + D_1(1) \\ A_1(q) &= D_1(q), \qquad q > 1. \end{aligned}$$

Subtracting (8.1) from (8.0) we have for any q

$$A(q) = D(q)$$

where $A(q)$ and $D(q)$ are the invariant points of T^q, but not T_k, $0 < k < q$, contained in J. This result, as with (5.3) in the proof of Theorem IV, leads to statement of Theorem V.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ANNALS OF MATHEMATICS
Vol. 4, No. 45, October, 1944

COMPLETENESS THEOREMS OF PALEY-WIENER TYPE

BY HARRY POLLARD

(Received December 14, 1943, revised January 14, 1944)

Various writers have obtained criteria for the completeness of a set $\{y_n\}$ in Hilbert space H, ([1], [3], [5, 100], [6]). These state that if $\{y_n\}$ is "near" a complete set $\{x_n\}$ in some suitable sense then $\{y_n\}$ is complete. In this paper we contribute a further result of this character, and discuss some corollaries.

1. The main theorem

Our principal criterion is the following.

THEOREM 1.1. *Let $\{x_n\}$, $\{y_n\}$ be sequences in H. Suppose that for fixed λ_1, λ_2, $0 \leqq \lambda_1 < 1$, $0 \leqq \lambda_2 < 1$ the following inequalities are satisfied by every finite sequence of numbers $\{a_n\}$:*

$$\|\Sigma a_n(x_n - y_n)\|^2 \leqq \lambda_1 \|\Sigma a_n x_n\|^2 + \lambda_2 \|\Sigma a_n y_n\|^2. \tag{1.1}$$

Then if $\{x_n\}$ is complete so is $\{y_n\}$.

For $\lambda_2 = 0$ this result is due to Paley and Wiener [5, 100], the form to Boas [2, 469]. It should be remarked that while our restrictions are less severe due to the presence of the λ_2 term, our conclusion is correspondingly weaker. For Paley and Wiener obtain also an expansion theory, and this we lack.

To prove our theorem we may suppose that $\lambda_1 = \lambda_2 = \lambda$, since we may replace λ_1 and λ_2 by $\lambda = \max(\lambda_1, \lambda_2) < 1$. We shall show that $(y_n, z) = 0$, $n = 1, 2, \cdots$ implies $z = 0$. Since $\{x_n\}$ is complete there exist coefficients a_{nk} such that

$$z = \lim_{k \to \infty} \sum_n a_{nk} x_n;$$

the sums are finite. Let $z_k = \sum_n a_{nk} x_n$, $w_k = \sum a_{nk} y_n$, noting that $(z, w_k) = 0$ for all k. By (1.1) and Schwarz's inequality

$$\begin{aligned}(1 - \lambda)(\|z_k\|^2 + \|w_k\|^2) &\leqq \|z_k\|^2 + \|w_k\|^2 - \|z_k - w_k\|^2 \\ &= 2R(z_k, w_k) = 2R(z_k - z, w_k) \leqq 2\|z_k - z\|\,\|w_k\|,\end{aligned}$$

whence

$$\|w_k\|^2 \leqq \|z_k\|^2 + \|w_k\|^2 \leqq \frac{2}{1 - \lambda}\|z_k - z\|\,\|w_k\|. \tag{1.2}$$

Then $0 \leqq \|w_k\| \leqq \dfrac{2}{1-\lambda}\|z_k - z\|$. Letting $k \to \infty$ it follows that $\|w_k\| \to 0$. Passage to the limit $(k \to \infty)$ in (1.2) yields $0 \leqq \|z\| \leqq 0$, whence $z = 0$.

2. Further criteria

Note that under the hypotheses of Theorem 1.1 the sets $\{x_n\}$, $\{y_n\}$ are both complete or both incomplete. The following criterion can also be deduced from the main theorem.

THEOREM 2.1. *If the inequalities*

$$\| \Sigma a_n(x_n - y_n) \| \leqq \lambda\{\| \Sigma a_n x_n \| + \| \Sigma a_n y_n \|\} \tag{2.1}$$

are satisfied by every finite sequence $\{a_n\}$ *for a fixed* $\lambda < \frac{1}{2}\sqrt{2}$, *then both sets are complete or both incomplete.*

PROOF: The inequality $(A + B)^2 \leqq 2(A^2 + B^2)$ shows that (2.1) implies

$$\| \Sigma a_n(x_n - y_n) \|^2 \leqq 2\lambda^2(\| \Sigma a_n x_n \|^2 + \| \Sigma a_n y_n \|^2).$$

If $2\lambda^2 < 1$ then (1.1) holds with $\lambda_1 = \lambda_2 = 2\lambda^2 < 1$; and the sets $\{x_n\}$ and $\{y_n\}$ are both complete or both incomplete.[1]

Of particular interest is the result obtained from Theorem 1.1 if we require that $\{x_n\}$ and $\{y_n\}$ both be orthonormal sets (denoted hereafter by ON).

THEOREM 2.2. *Let* $\{x_n\}$, $\{y_n\}$ *be two ON sets in H such that the inequalities*

$$\Sigma\Sigma a_m \bar{a}_n (x_m, y_n) \geqq \mu \Sigma \mid a_n \mid^2$$

hold for all finite sets $\{a_n\}$ *and a fixed* $\mu > 0$. *Then both are complete or both incomplete.*

3. The bounds on λ_1, λ_2

It is clear that the value $\lambda_1 = 1$ is not admissible in Theorem 1.1; for let $\{y_n\}$ be the sequence $\{0, 0, \cdots\}$.

Now let z_n be any complete ON set. Define $x_n = z_{n+1} - z_n$, $y_n = z_{n+1}$. Then $\{x_n\}$ is complete, $\{y_n\}$ incomplete, and

$$\| \Sigma a_n(x_n - y_n) \|^2 \leqq \| \Sigma a_n y_n \|^2. \tag{3.1}$$

This proves that $\lambda_2 = 1$ is inadmissible. (This example was suggested by a similar one in [3, 852–3]).

The following questions remain open. (i) Is it true that for every complete set $\{x_n\}$ there exists an incomplete set $\{y_n\}$ such that (3.1) is true? (ii) Are the bounds on λ and μ in Theorems 2.1 and 2.2 best possible?

KENYON COLLEGE

REFERENCES

1. BARY, N., *Sur la stabilité de la propriété d'être un système complet de fonctions*, Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS, vol. 37 (1942), pp. 83–87.
2. BOAS, R. P., JR., *Expansions of analytic functions*, Transactions of the American Mathematical Society, vol. 48 (1940), pp. 467–487.
3. DUFFIN AND EACHUS, *Some notes on an expansion theorem of Paley and Wiener*, Bulletin of the American Mathematical Society, vol. 48 (1942), pp. 850–855.
4. KACMARCZ AND STEINHAUS, *Orthogonalreihen*, Warsaw, 1935.
5. PALEY AND WIENER, *Fourier Transforms in the Complex Domain*, New York, 1934.
6. TSENG, Y. Y., *On generalized biorthogonal expansions in metric and unitary spaces*, Proceedings of the National Academy of Sciences, vol. 28 (1942), pp. 170–175.

[1] My original theorem required $\lambda < \frac{1}{2}$. The improvement is due to the referee.

ANNALS OF MATHEMATICS
Vol. 45, No. 4, October, 1944

THE ANALOGUE OF BROMWICH'S THEOREM FOR INTEGRAL TRANSFORMATIONS

BY H. L. GARABEDIAN

(Received December 23, 1943; Revised April 7, 1944)

1. Introduction

In essence, Bromwich's theorem for matrix transformations [1] provides sufficient conditions in order that a method of summation with infinite matrix of reference shall include Cesàro summability, a method of summation with finite matrix of reference. Various additional proofs of this theorem may be found in the mathematical literature, the most recent one by Garabedian [2], and also a number of applications, notably by Moore [3], Morse [4], and Garabedian [5].

This paper involves transformations of the type

$$\sigma(x) = \int_0^\infty K(x, t)s(t)\,dt, \tag{1.1}$$

which associate with certain functions $s(t)$ the functions $\sigma(t)$ determined by a given *kernel* $K(x, t)$. Restrictions on $s(t)$ and $K(x, t)$ for the present investigation will be specified later in this section. We shall also be concerned with transformations of the type

$$\sigma(x) = \int_0^x K(x, t)s(t)\,dt, \tag{1.2}$$

obtainable directly from (1.1) in the case that $K(x, t) \equiv 0$, $t > x$. The transformations (1.1) and (1.2) (or the kernels in (1.1) and (1.2)) are said to be of *infinite reference* and *finite reference* respectively. An integral transformation or its kernel is said to be *regular* if $\lim_{x\to\infty} s(x)$ implies the existence of $\lim_{x\to\infty} \sigma(x)$ and the equality of the two limits.

In this paper it will be of convenience also to use transformations of the type

$$z(x) = \int_0^\infty k(x, t)a(t)\,dt, \tag{1.3}$$

where $s(t) = \int_0^t a(u)\,du$.

We follow the procedure of Knopp [6] in setting down restrictions on $s(t)$ and $K(x, t)$.

Let S be the class of functions which satisfy the following three conditions:

(i) *$s(t)$ is defined for $t > 0$,*
(ii) *$s(t)$ is bounded for $0 < t \leqq a$, for every $a > 0$,*
(iii) *$s(t)$ is Lebesgue integrable on the interval $(0, a)$, for every $a > 0$.*

We designate by S_b the subclass of functions in S for which $s(t)$ is bounded, $t > 0$.

Let $K(x, t)$ be defined and measurable in the quadrant Q: $(x > 0, t > 0)$. Then we require that

(i) $\int_0^\infty | K(x, t) | \, dt$ *shall exist for almost all* x *and belong to* S_b; in particular

$$\int_0^\infty | K(x, t) | \, dt < M,$$

M independent of x,

$$\lim_{x\to\infty} \int_0^a | K(x, t) | \, dt = 0, \tag{1.4)(ii}$$

for every fixed $a > 0$,

$$\lim_{x\to\infty} \int_0^\infty K(x, t) \, dt = 1. \tag{iii}$$

Knopp [6] has proved that the conditions (1.4) ensure the regularity of the transformation (1.1). In this connection it is understood that $s(t)$ belongs to S. For this paper we do not require conditions of regularity for transformations of the type (1.3).

The regular transformation

$$y(t) = \int_0^t \left(1 - \frac{u}{t}\right)^n a(u) \, du, \tag{1.5}$$

which defines Cesàro summability (C, n), $n > 0$, for integral transforms, is of particular interest in this paper.

It is the main object of this paper to establish the analogue of Bromwich's theorem for integral transformations, that is, to provide sufficient conditions in order that a method of summation defined by a transformation of the type (1.3) shall include Cesàro summability of positive integral order. The proof is given in §2. Bromwich [1] has established a theorem resembling the theorem of §2, but only for the case of $(C, 1)$ summability and with less generality and simplicity than achieved by this writer. In §3 a kernel $k(x, t)$ which fulfills the conditions of the theorem of §2 is exhibited. In §4 the connection of the Bromwich theorem for integral transformations with the problem of the flow of heat in an infinite rod whose surface is a non-conductor is discussed.

2. The analogue of the Bromwich theorem for integral transformations

This section is devoted to the statement and the proof of the analogue of the Bromwich theorem for integral transformations.

THEOREM 2.1. *Let the integral* $\int_0^\infty a(u) \, du$, $a(u)$ *in* S, *be summable* (C, n), *for some positive integral value of* n, *to the value* l. *Let* $k(x, t)$, *together with its first*

$n + 1$ derivatives, be defined and measurable in the quadrant Q: $(x > 0, t > 0)$. Let $K(x, t) = (-1)^{n+1}t^n(\partial^{n+1}/\partial t^{n+1})k(x, t)/\Gamma(n + 1)$ and require that

(i) $\int_0^\infty | K(x, t) | dt$ *shall exist for almost all x and belong to S_b; in particular,*

$$\int_0^\infty | K(x, t) | dt < M,$$

M independent of x,

(ii) $$\lim_{x\to\infty} \int_0^a | K(x, t) | dt = 0,$$

for every fixed $a > 0$,

(iii) $$\lim_{x\to\infty} \int_0^\infty K(x, t)\, dt = 1,$$

(iv) $$\lim_{t\to\infty} t^i \frac{\partial^i}{\partial t^i} k(x, t) = 0, \qquad i = 0, 1, 2, \cdots, n; \quad x \geqq \delta > 0.$$

Then the integral

$$\int_0^\infty k(x, t)a(t)\, dt$$

exists for almost all x, and

$$\lim_{x\to\infty} \int_0^\infty k(x, t)a(t)\, dt = l.$$

We are concerned with the Cesàro transformation

$$t^n y(t) = \int_0^t (t - u)^n a(u)\, du, \tag{2.1}$$

and the transformation

$$z(x) = \int_0^\infty k(x, t)a(t)\, dt. \tag{2.2}$$

We wish to determine conditions on the kernel $k(x, t)$ so that $\lim_{x\to\infty} y(x) = l$ shall imply $\lim_{x\to\infty} z(x) = l$. In this connection the problem of expressing z in terms of y presents itself. We do this formally at first, making all necessary assumptions to determine the *form* of the required transformation. Our method of formal procedure will be to solve (2.1) for a in terms of y and then substitute in (2.2), obtaining z in terms of y.

To solve (2.1) for a in terms of y we observe first of all that the expression $\int_0^t (t - u)^n a(u)\, du/\Gamma(n + 1)$ is the $(n + 1)^{st}$ integral of $a(u)$ in the Liouville-

Riemann theory of fractional integration. Accordingly, we have

$$D_t^{n+1}\{t^n y(t)\} = \Gamma(n+1)a(t), \tag{2.3}$$

this formula being valid for almost all t.

Now, eliminating a between (2.2) and (2.3), we have

$$z(x) = \frac{1}{\Gamma(n+1)} \int_0^\infty k(x, t) D_t^{n+1}\{t^n y(t)\}\, dt.$$

Integrating by parts in this equation we get

$$z(x) = \frac{-1}{\Gamma(n+1)} \int_0^\infty \frac{\partial}{\partial t} k(x, t) D_t^n \{t^n y(t)\}\, dt,$$

on the assumption that $k(x, t)D_t^n\{t^n y(t)\}$ vanishes as $t \to 0$ and as $t \to \infty$. Finally, after other assumptions of this nature and after n additional integrations by parts, we obtain

$$z(x) = \frac{(-1)^{n+1}}{\Gamma(n+1)} \int_0^\infty t^n \frac{\partial^{n+1}}{\partial t^{n+1}} k(x, t)\, dt. \tag{2.4}$$

It is clear that this is not a valid transformation from y to z unless we impose restrictions of unwarranted severity on the class of functions a which we are considering. Accordingly, we shall look for a means of obtaining (2.4) directly without further limiting the class of functions a in S. To this end we eliminate y between (2.1) and (2.4) and attempt to derive (2.2) without adding to the original restrictions on a. Thus, we have at the outset

$$z(x) = \frac{(-1)^{n+1}}{\Gamma(n+1)} \int_0^\infty \frac{\partial^{n+1}}{\partial t^{n+1}} k(x, t) \int_0^t (t-u)^n a(u)\, du\, dt. \tag{2.5}$$

The integral (2.4), and consequently (2.5), exists by virtue of condition (i) of Theorem 2.1 and the boundedness and integrability of y. Then, we enlist the aid of the theorem of Fubini [7] to justify an interchange in the order of integration in (2.5) to obtain

$$z(x) = \frac{(-1)^{n+1}}{\Gamma(n+1)} \int_0^\infty a(u) \int_u^\infty (t-u)^n \frac{\partial^{n+1}}{\partial t^{n+1}} k(x, t)\, dt\, du. \tag{2.6}$$

We wish now to prove that

$$\frac{(-1)^{n+1}}{\Gamma(n+1)} \int_u^\infty (t-u)^n \frac{\partial^{n+1}}{\partial t^{n+1}} k(x, t)\, dt = k(x, u). \tag{2.7}$$

Using condition (iv) of Theorem 2.1 and integrating by parts n times in the left member of (2.7) we get

$$\frac{(-1)^{n+1}}{\Gamma(n+1)} \int_u^\infty (t-u)^n \frac{\partial^{n+1}}{\partial t^{n+1}} k(x, t)\, dt = -\int_u^\infty \frac{\partial}{\partial t} k(x, t)\, dt = k(x, u).$$

From this relationship and (2.6) we have

$$z(x) = \int_0^\infty k(x, u)a(u)\, du, \tag{2.8}$$

which is identical with (2.2) and which we have shown to exist for almost all x. This establishes the transformation (2.4) from z to y without change of our original hypotheses.

It remains to prove that $\lim_{x\to\infty} z(x) = l$. To this end we need merely to require that (2.4) be a regular transformation, since $\lim_{x\to\infty} y(x) = l$ by assumption. The requirements (1.4) for regularity of a transformation of the type (2.4) are fulfilled by virtue of conditions (i), (ii), and (iii) of our theorem. Our theorem is then established.

3. A kernel which fulfills the conditions of Theorem 2.1

In this section we show that the kernel $k(x, t) = e^{-t^2/x}$ fulfills the conditions of Theorem 2.1. This kernel when associated with the transformation (1.3) generates an integral transformation which is the analogue of one of the Dirichlet's series transformations [5] in the field of matrix transformations. The kernel $k(x, t) = e^{-t/x}$ provides the analogue of the Abel transformation in the field of matrix transformations. It can be shown that the Abel kernel also fulfills the conditions of Theorem 2.1, but the proof is omitted here.

In order to check the conditions of Theorem 2.1 for the kernel $k(x, t) = e^{-t^2/x}$ we need the formula

$$D_z^n e^{-z^2} = (-1)^n H_n(z)e^{-z^2},$$

where $H_n(z)$ is the n^{th} Hermite polynomial. Thus, we have

$$\frac{\partial^n}{\partial t^n} e^{-t^2/x} = (-1)^n x^{-n/2} H_n(tx^{-1/2})e^{-t^2/x}, \tag{3.1}$$

and consequently

$$K(x, t) = \frac{(-1)^{n+1} t^n}{n!} \frac{\partial^{n+1}}{\partial t^{n+1}} k(x, t) = \frac{1}{n!} (tx^{-1/2})^n H_{n+1}(tx^{-1/2})e^{-t^2/x} x^{-1/2}.$$

Now, we have

$$\int_0^a | K(x, t) |\, dt = \frac{1}{n!} \int_0^{ax^{-1/2}} u^n | H_{n+1}(u) | e^{-u^2}\, du, \tag{3.2}$$

whence we conclude that $\lim_{x\to\infty} \int_0^a | K(x, t) |\, dt = 0$. Accordingly, condition (ii) of Theorem 2.1 is fulfilled.

From (3.2) we have

$$\int_0^\infty | K(x, t) |\, dt = \frac{1}{n!} \int_0^\infty u^n | H_{n+1}(u) | e^{-u^2}\, du.$$

Now, $|H_{n+1}(u)| \leqq P_{n+1}(u)$, where $P_{n+1}(u)$ is $H_{n+1}(u)$ with all coefficients positive. Since $\int_0^\infty u^n P_{n+1}(u)e^{-u^2}\,du$ exists and is independent of x, we conclude that $\int_0^\infty |K(x, t)|\,dt < M$, M independent of x, and hence that condition (i) is satisfied.

With the aid of (3.1) it is clear that condition (iv) is fulfilled.

Finally, we consider

$$\int_0^\infty K(x, t)\,dt = \frac{1}{n!}\int_0^\infty u^n H_{n+1}(u)e^{-u^2}\,du = \frac{(-1)^{n+1}}{n!}\int_0^\infty u^n D_u^{n+1} e^{-u^2}\,du,$$

and observe that n integrations by parts transforms the last integral into $-\int_0^\infty d(e^{-u^2}) = 1$. This proves that (iii) holds and completes the discussion of this section.

4. Application to the rigorous solution of the problem of the flow of heat in an infinite rod

We consider an infinite rod whose surface is a non-conductor, and we suppose that any cross section of the bar is so small that the temperature is constant in that cross section. We suppose further that the initial distribution of the temperature is given by the function $f(x)$. It is required to determine the temperature of any point of the bar at any later instant. The differential equation dominating the unsteady flow of heat in one dimension is

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2}, \qquad t > 0, \tag{4.1}$$

where $u(x, t)$ is the temperature function. We seek a solution of the differential equation subject to the condition.

$$\lim_{t \to 0} u(x, t) = f(x), \tag{4.2}$$

provided that $f(x)$ is continuous. If we assume simply that $f(x)$ is Lebesgue integrable on the interval $(-\infty, \infty)$, the limit in (4.2) is equal to $\frac{1}{2}[f(x + 0) + f(x - 0)]$ wherever this expression has a meaning; to $f(x)$ wherever $f(x)$ is continuous; and to $f(x)$ for almost all values of x.

Using a classical formal procedure we find that the integral

$$\int_0^\infty \left[A(\beta) \cos \frac{\beta x}{a} + B(\beta) \sin \frac{\beta x}{a} \right] e^{-\beta^2 t}\, d\beta \tag{4.3}$$

may afford a solution of the differential equation. This integral must tend to $f(x)$, if $f(x)$ is continuous, as $t \to 0$. This raises the question of the possibility of expressing $f(x)$ in the form

$$\int_0^\infty \left[A(\beta) \cos \frac{\beta x}{a} + B(\beta) \sin \frac{\beta x}{a} \right] d\beta. \tag{4.4}$$

The formal development of $f(x)$ in terms of a Fourier integral is given by (4.4) where

$$A(\beta) = \frac{1}{\pi a}\int_{-\infty}^{\infty} f(\lambda)\cos\frac{\beta\lambda}{a}\,d\lambda, \tag{4.5}$$
$$B(\beta) = \frac{1}{\pi a}\int_{-\infty}^{\infty} f(\lambda)\sin\frac{\beta\lambda}{a}\,d\lambda.$$

Substitution of A and B, as given in (4.5), into (4.3) gives the expression

$$\frac{1}{\pi a}\int_0^{\infty}\int_{-\infty}^{\infty} f(\lambda)\cos\frac{\beta}{a}(\lambda - x)e^{-\beta^2 t}\,d\lambda\,d\beta. \tag{4.6}$$

Integration with respect to β in (4.6) identifies this integral with the classical solution of Poisson to the proposed problem.

It is easy to show that the integral (4.6) converges for $t \geqq t_0 > 0$, and all values of x, and defines in that region a continuous function $u(x, t)$ which satisfies the differential equation (4.1). In order to satisfy condition (4.2) we have merely to show that the kernel $k(1/t, \beta) = e^{-\beta^2 t}$ satisfies the conditions of Theorem 2.1, since, by Fejér's theorem [8, p. 29], the integral (4.4) is summable $(C, 1)$ to $\frac{1}{2}[f(x+0) + f(x-0)]$ wherever this expression has a meaning; to $f(x)$ wherever $f(x)$ is continuous; and to $f(x)$ for almost all values of x. Since, in §3, we proved that the kernel $k(1/t, \beta) = e^{-\beta^2 t}$ fulfills the conditions of Theorem 2.1 we conclude that $\lim_{t\to 0} u(x, t)$ is equal to $\frac{1}{2}[f(x+0) + f(x-0)]$ wherever this expression has a meaning; to $f(x)$ wherever $f(x)$ is continuous; and to $f(x)$ for almost all values of x.

The problem under discussion in this section affords an interesting application of Theorem 2.1. It should, however, be noted that the solution is valid under much less restrictive conditions on $f(x)$ at infinity than imposed by this writer [8, p. 31].

NORTHWESTERN UNIVERSITY

BIBLIOGRAPHY

1. T. J. BROMWICH, *On the limits of certain infinite series and integrals*, Mathematische Annalen, vol. 65 (1907–08), pp. 350–369.
2. H. L. GARABEDIAN, *Note on a theorem due to Bromwich*, Bulletin of the American Mathematical Society, vol. 38 (1932), pp. 541–545.
C. N. MOORE, *Applications of the theory of summability to developments in orthogonal functions*, Bulletin of the American Mathematical Society, vol. 25 (1918–19), pp. 258–276.
4. D. S. MORSE, *Relative inclusiveness of certain definitions of summability*, American Journal of Mathematics, vol. 45 (1923), pp. 259–285.
5. H. L. GARABEDIAN, *On the relation between certain methods of summability*, Annals of Mathematics, vol. 32 (1931), pp. 83–106.
6. K. KNOPP, *Über eine Klasse konvergenzerhaltenden Integraltransformationen und den Äquivalenzsatz der C- und H- Verfahren*, Mathematische Zeitschrift, vol. 47 (1941), pp. 229–264.
7. C. CARATHÉODORY, *Vorlesungen über reele Funktionen*, Leipzig und Berlin, 1918.
8. E. C. TITCHMARSH, *Introduction to the theory of Fourier integrals*, Oxford University Press, 1937.

Annals of Mathematics
Vol. 45, No. 4, October, 1944

A SIMPLE INTRINSIC PROOF OF THE GAUSS-BONNET FORMULA FOR CLOSED RIEMANNIAN MANIFOLDS

By Shiing-shen Chern

(Received November 26, 1943)

Introduction

C. B. Allendoerfer[1] and W. Fenchel[2] have independently given a generalization of the classical formula of Gauss-Bonnet to a closed orientable Riemannian manifold which can be imbedded in a euclidean space. Recently, Allendoerfer and André Weil[3] extended the formula to a closed Riemannian polyhedron and proved in particular its validity in the case of a general closed Riemannian manifold. In their proof use is still made of the imbedding of a Riemannian cell in a euclidean space. The object of this paper is to offer a direct intrinsic proof of the formula by making use of the theory of vector fields in differentiable manifolds.

The underlying idea of the present proof is very simple, so that a brief summary might be helpful. Let R^n be a closed orientable Riemannian manifold of an even dimension n. According to details to be given below, we define in R^n an intrinsic exterior differential form Ω of degree n, which is of course equal to a scalar invariant of R^n multiplied by the volume element. The formula of Gauss-Bonnet in question asserts that the integral of this differential form over R^n is equal to the Euler-Poincaré characteristic χ of R^n. To prove this we pass from the manifold R^n to the manifold M^{2n-1} of $2n - 1$ dimensions formed by the unit vectors of R^n.[4] In M^{2n-1} we show that Ω is equal to the exterior derivative of a differential form Π of degree $n - 1$. By defining a continuous field of unit vectors over R^n with isolated singular points, we get, as its image in M^{2n-1}, a submanifold V^n of dimension n, and the integral of Ω over R^n is equal to the same integral over V^n. The application of the theorem of Stokes shows that the latter is equal to the integral of Π over the boundary of V^n. Now, the boundary of V^n corresponds exactly to the singular points of the vector field defined in R^n, the sum of whose indices is, by a well-known theorem, equal to χ. With such an interpretation the integral of Π over the boundary of V^n can be evaluated and is easily proved to be equal to χ.

The method can of course be applied to derive other formulas of the same type and, with suitable modifications, to deduce the Gauss-Bonnet formula for a Riemannian polyhedron. We publish this proof, because it is in the present case that the main ideas of our method are most clear. Further results will be given in a forthcoming paper.

§1. Résumé of some fundamental formulas in Riemannian Geometry

Let R^n be a closed orientable differentiable manifold[5] of an even dimension $n = 2p$ and class $r \geqq 4$. In R^n suppose a Riemannian metric be defined, with

the fundamental tensor g_{ij}, whose components we suppose to be of class 3. Since we are to deal with multiple integrals, it seems convenient to follow Cartan's treatment of Riemannian Geometry,[6] with the theory of exterior differential forms, instead of the ordinary tensor analysis, playing the dominant rôle. The differential forms which occur below are exterior differential forms.

According to Cartan we attach to each point P of R^n a set of n mutually perpendicular unit vectors $\mathfrak{e}_1, \cdots, \mathfrak{e}_n$, with a certain orientation. Such a figure $P\mathfrak{e}_1 \cdots \mathfrak{e}_n$ is called a frame. A vector $\mathfrak{v}$ of the tangent space of R^n at P can be referred to the frame at P, thus

$$\mathfrak{v} = u_i \mathfrak{e}_i, \tag{1}$$

where the index i runs from 1 to n and repeated indices imply summation. The law of infinitesimal displacement of tangent spaces, as defined by the parallelism of Levi-Civita, is given by equations of the form

$$\begin{cases} dP = \omega_i \mathfrak{e}_i, \\ d\mathfrak{e}_i = \omega_{ij} \mathfrak{e}_j, \qquad \omega_{ij} + \omega_{ji} = 0 \end{cases} \tag{2}$$

where ω_i, ω_{ij} are Pfaffian forms. These Pfaffian forms satisfy the following "equations of structure":

$$\begin{cases} d\omega_i = \omega_j \omega_{ji}, \\ d\omega_{ij} = -\omega_{ik}\omega_{jk} + \Omega_{ij}, \qquad \Omega_{ij} + \Omega_{ji} = 0. \end{cases} \tag{3}$$

In (3) Ω_{ij} are exterior quadratic differential forms and give the curvature properties of the space.

The forms Ω_{ij} satisfy a system of equations obtained by applying to (3) the theorem that the exterior derivatives of the left-hand members are zero. The equations are

$$\begin{cases} \omega_j \Omega_{ji} = 0, \\ d\Omega_{ij} - \omega_{jk}\Omega_{ik} + \omega_{ik}\Omega_{jk} = 0, \end{cases} \tag{4}$$

and are called the Bianchi identities.

For the following it is useful to know how the Ω_{ij} behave when the frame $\mathfrak{e}_1 \cdots \mathfrak{e}_n$ undergoes a proper orthogonal transformation. In a neighborhood of P in which the same system of coordinates is valid let $\mathfrak{e}_1 \cdots \mathfrak{e}_n$ be changed to $\mathfrak{e}_1^* \cdots \mathfrak{e}_n^*$ according to the proper orthogonal transformation:

$$\mathfrak{e}_i^* = a_{ij}\mathfrak{e}_j \tag{5}$$

or

$$\mathfrak{e}_i = a_{ji}\mathfrak{e}_j^*, \tag{5'}$$

where (a_{ij}) is a proper orthogonal matrix, whose elements a_{ij} are functions of the coordinates. Suppose Ω_{ij}^* be formed from the frames $P\mathfrak{e}_1^* \cdots \mathfrak{e}_n^*$ in the same way as Ω_{ij} are formed from $P\mathfrak{e}_1 \cdots \mathfrak{e}_n$. Then we easily find

$$\Omega_{ij}^* = a_{ik}a_{jl}\Omega_{kl}. \tag{6}$$

From (6) we deduce an immediate consequence. Let $\epsilon_{i_1 \cdots i_n}$ be a symbol which is equal to $+1$ or -1 according as $i_1, \cdots, i_n$ form an even or odd permutation of $1, \cdots, n$, and is otherwise zero. Since our space R^n is of even dimension $n = 2p$, we can construct the sum

$$(7) \qquad \Omega = (-1)^{p-1} \frac{1}{2^{2p} \pi^p p!} \epsilon_{i_1 \cdots i_{2p}} \Omega_{i_1 i_2} \Omega_{i_3 i_4} \cdots \Omega_{i_{2p-1} i_{2p}},$$

where each index runs from 1 to n. Using (6), we see that Ω remains invariant under a change of frame (5) and is therefore intrinsic. This intrinsic differential form Ω is of degree n and is thus a multiple of $\omega_1 \cdots \omega_n$. As the latter product (being the volume element of the space) is also intrinsic, we can write

$$(8) \qquad \Omega = I\omega_1 \cdots \omega_n,$$

where the coefficient I is a scalar invariant of the Riemannian manifold.

With all these preparations we shall write the formula of Gauss-Bonnet in the following form

$$(9) \qquad \int_{R^n} \Omega = \chi,$$

χ being the Euler-Poincaré characteristic of R^n.

§2. The space of unit vectors and a formula for Ω

From the Riemannian manifold R^n we pass now to the manifold M^{2n-1} of dimension $2n - 1$ formed by its unit vectors. M^{2n-1} is a closed differentiable manifold of class $r - 1$. As its local coordinates we may of course take the local coordinates of R^n and the components u_i of the vector $\mathfrak{v}$ in (1), subjected to the condition

$$(1') \qquad u_i u_i = 1.$$

If θ_i are the components of $d\mathfrak{v}$ with respect to the frame $\mathfrak{e}_1 \cdots \mathfrak{e}_n$, we have

$$(10) \qquad d\mathfrak{v} = \theta_i \mathfrak{e}_i,$$

where

$$(11) \qquad \theta_i = du_i + u_j \omega_{ji}$$

and

$$(12) \qquad u_i \theta_i = 0.$$

From (11) we get, by differentiation,

$$(13) \qquad d\theta_i = \theta_j \omega_{ji} + u_j \Omega_{ji}.$$

As to the effect of a change of frame (5) on the components u_i, θ_i, it is evidently given by the equations

$$(14) \qquad u_i^* = a_{ij} u_j, \qquad \theta_i^* = a_{ij} \theta_j.$$

We now construct the following two sets of differential forms:

$$(15)\qquad \Phi_k = \epsilon_{i_1 \cdots i_{2p}} u_{i_1} \theta_{i_2} \cdots \theta_{i_{2p-2k}} \Omega_{i_{2p-2k+1} i_{2p-2k+2}} \cdots \Omega_{i_{2p-1} i_{2p}}, \qquad k = 0, 1, \cdots, p-1,$$

$$(16)\qquad \Psi_k = \epsilon_{i_1 \cdots i_{2p}} \Omega_{i_1 i_2} \theta_{i_3} \cdots \theta_{i_{2p-2k}} \Omega_{i_{2p-2k+1} i_{2p-2k+2}} \cdots \Omega_{i_{2p-1} i_{2p}}, \qquad k = 0, 1, \cdots, p-1.$$

The forms Φ_k are of degree $2p - 1$ and Ψ_k of degree $2p$, and we remark that Ψ_{p-1} differs from Ω only by a numerical factor. Using (6) and (14), we see that Φ_k and Ψ_k are intrinsic and are therefore defined over the entire Riemannian manifold R^n.

We shall prove the following recurrent relation:

$$(17)\qquad d\Phi_k = \Psi_{k-1} + \frac{2p - 2k - 1}{2(k+1)} \Psi_k, \qquad k = 0, 1, \cdots, p-1,$$

where we define $\Psi_{-1} = 0$. Using the property of skew-symmetry of the symbol $\epsilon_{i_1 \cdots i_{2p}}$ in its indices, we can write

$$\begin{aligned} d\Phi_k = {} & \epsilon_{(i)} du_{i_1} \theta_{i_2} \cdots \theta_{i_{2p-2k}} \Omega_{i_{2p-2k+1} i_{2p-2k+2}} \cdots \Omega_{i_{2p-1} i_{2p}} \\ & + (2p - 2k - 1) \epsilon_{(i)} u_{i_1} d\theta_{i_2} \theta_{i_3} \cdots \theta_{i_{2p-2k}} \Omega_{i_{2p-2k+1} i_{2p-2k+2}} \cdots \Omega_{i_{2p-1} i_{2p}} \\ & - k \epsilon_{(i)} u_{i_1} \theta_{i_2} \cdots \theta_{i_{2p-2k}} d\Omega_{i_{2p-2k+1} i_{2p-2k+2}} \Omega_{i_{2p-2k+3} i_{2p-2k+4}} \cdots \Omega_{i_{2p-1} i_{2p}}, \end{aligned}$$

where $\epsilon_{(i)}$ is an abbreviation of $\epsilon_{i_1 \cdots i_{2p}}$. For the derivatives du_i, $d\theta_i$, $d\Omega_{ij}$ we can substitute their expressions from (11), (13), and (4). The resulting expression for $d\Phi_k$ will then consist of terms of two kinds, those involving ω_{ij} and those not. We collect the terms not involving ω_{ij}, which are

$$(18)\qquad \Psi_{k-1} + (2p - 2k - 1) \epsilon_{(i)} u_{i_1} u_j \Omega_{j i_2} \theta_{i_3} \cdots \theta_{i_{2p-2k}} \Omega_{i_{2p-2k+1} i_{2p-2k+2}} \cdots \Omega_{i_{2p-1} i_{2p}}.$$

This expression is obviously intrinsic. Its difference with $d\Phi_k$ is an expression which contains a factor ω_{ij} in each of its terms.

We shall show that this difference is zero. In fact, let P be an arbitrary but fixed point of R^n. In a neighborhood of P we can choose a family of frames $e_1 \cdots e_n$ such that at P,

$$\omega_{ij} = 0.$$

(This process is "equivalent" to the use of geodesic coordinates in tensor notation.) Hence, for this particular family of frames, the expressions (18) and $d\Phi_k$ are equal at P. It follows that they are identical, since both expressions are intrinsic and the point P is arbitrary.

To transform the expression (18) we shall introduce the abbreviations

$$(19)\qquad \begin{cases} P_k = \epsilon_{(i)} u_{i_1}^2 \Omega_{i_1 i_2} \theta_{i_3} \cdots \theta_{i_{2p-2k}} \Omega_{i_{2p-2k+1} i_{2p-2k+2}} \cdots \Omega_{i_{2p-1} i_{2p}}, \\ \Sigma_k = \epsilon_{(i)} u_{i_1} u_{i_3} \Omega_{i_3 i_2} \theta_{i_3} \cdots \theta_{i_{2p-2k}} \Omega_{i_{2p-2k+1} i_{2p-2k+2}} \cdots \Omega_{i_{2p-1} i_{2p}}, \\ T_k = \epsilon_{(i)} u_{i_3}^2 \Omega_{i_1 i_2} \theta_{i_3} \cdots \theta_{i_{2p-2k}} \Omega_{i_{2p-2k+1} i_{2p-2k+2}} \cdots \Omega_{i_{2p-1} i_{2p}}, \end{cases}$$

which are forms of degree $2p$. Owing to the relations (1′) and (12) there are some simple relations between these forms and Ψ_k. In fact, we can write

$$P_k = \epsilon_{(i)}(1 - u_{i_2}^2 - u_{i_3}^2 - \cdots - u_{i_{2p}}^2)\Omega_{i_1 i_2}\theta_{i_3} \cdots \theta_{i_{2p-2k}}\Omega_{i_{2p-2k+1}i_{2p-2k+2}} \cdots \Omega_{i_{2p-1}i_{2p}}$$
$$= \Psi_k - P_k - 2(p - k - 1)T_k - 2kP_k,$$

which gives

$$\Psi_k = 2(k+1)P_k + 2(p-k-1)T_k. \tag{20}$$

Again, we have

$$\Sigma_k = \epsilon_{(i)} u_{i_1}\Omega_{i_3 i_2}(-u_{i_1}\theta_{i_1} - u_{i_2}\theta_{i_2} - u_{i_4}\theta_{i_4} - \cdots - u_{i_{2p}}\theta_{i_{2p}})\theta_{i_4} \cdots$$
$$\theta_{i_{2p-2k}}\Omega_{i_{2p-2k+1}i_{2p-2k+2}} \cdots \Omega_{i_{2p-1}i_{2p}}$$
$$= T_k - (2k+1)\Sigma_k,$$

and hence

$$T_k = 2(k+1)\Sigma_k. \tag{21}$$

The expression (18) for $d\Phi_k$ therefore becomes

$$d\Phi_k = \Psi_{k-1} + (2p - 2k - 1)\{P_k + 2(p - k - 1)\Sigma_k\}, \qquad k = 0, 1, \cdots, p-1.$$

Using (20) and (21), we get the desired formula (17).

From (17) we can solve Ψ_k in terms of $d\Phi_0, d\Phi_1, \cdots, d\Phi_k$. The result is easily found to be

$$\psi_k = \sum_{m=0}^{k} (-1)^m \frac{2^{m+1}(k+1)k \cdots (k-m+1)}{(2p-2k-1)(2p-2k+1)\cdots(2p-2k+2m-1)} d\Phi_{k-m}, \qquad k = 0, 1, \cdots, p-1. \tag{22}$$

In particular, it follows that Ω is the exterior derivative of a form Π:

$$\Omega = (-1)^{p-1}\frac{1}{2^{2p}\pi^p p!}\Psi_{p-1} = d\Pi, \tag{23}$$

where

$$\Pi = \frac{1}{\pi^p}\sum_{m=0}^{p-1}(-1)^m \frac{1}{1 \cdot 3 \cdots (2p-2m-1)m!\,2^{p+m}}\Phi_m. \tag{24}$$

§3. Proof of the Gauss-Bonnet formula

Basing on the formula (24) we shall give a proof of the formula (9), under the assumption that R^n is a closed orientable Riemannian manifold.

We define in R^n a continuous field of unit vectors with a point 0 of R^n as the only singular point.[7] By a well-known theorem the index of the field at 0 is equal to χ, the Euler-Poincaré characteristic of R^n. This vector field defines in M^{2n-1} a submanifold V^n, which has as boundary χZ, where Z is the $(n-1)$-

dimensional cycle formed by all the unit vectors through 0. The integral of Ω over R^n is evidently equal to the same over V^n. Applying Stokes's theorem, we get therefore

$$(25)\qquad \int_{R^n} \Omega = \int_{V^n} \Omega = \chi \int_Z \Pi = \chi \frac{1}{1\cdot 3 \cdots (2p-1)2^p \pi^p} \int_Z \Phi_0 .$$

From the definition of Φ_0 we have

$$(26)\qquad \Phi_0 = (2p-1)! \sum_{i=1}^{n} (-1)^i \theta_1 \cdots \theta_{i-1} u_i \theta_{i+1} \cdots \theta_{2p} .$$

The last sum is evidently the volume element of the $(2p-1)$-dimensional unit sphere. Therefore

$$\int_Z \Phi_0 = (2p-1)! \frac{2\pi^p}{(p-1)!} .$$

Substituting this into (25), we get the formula (9).

INSTITUTE FOR ADVANCED STUDY, PRINCETON, N. J. AND
TSING HUA UNIVERSITY, KUNMING, CHINA.

REFERENCES

1. ALLENDOERFER, C. B., *The Euler number of a Riemann manifold*, Amer. J. Math., 62 (1940), 243–248.
2. FENCHEL, W., *On total curvatures of Riemannian manifolds I*, Jour. London Math. Soc., 15 (1940), 15–22.
3. ALLENDOERFER, C. B., AND ANDRÉ WEIL, *The Gauss-Bonnet theorem for Riemannian polyhedra*, Trans. Amer. Math. Soc., 53 (1943), 101–129.
4. For its definition and topology see, for instance, E. STIEFEL, *Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten*, Comm. Math. Helv., 8 (1936), 3–51.
5. WHITNEY, H., *Differentiable manifolds*, Annals of Math., 37 (1936), 645–680.
6. CARTAN, E., *Leçons sur la géomètrie des espaces de Riemann*, Paris 1928.
7. ALEXANDROFF-HOPF, Topologie I, 550.

Annals of Mathematics
Vol. 45, No. 4, October, 1944

THE TOPOLOGICAL THEORY OF FRÉCHET SURFACES[1]

By J. W. T. Youngs

(Received March 13, 1944)

Chapter I

Introduction

1. The term surface

The object of this paper is the study of surfaces. As the title indicates we do not propose to consider all the mathematical entities to which the term surface has been applied, and, in fact, our first obligation is to define precisely the objects of our attention.

In this connection even the most casual reader will have noticed a certain degree of confusion attached to the word surface as it is sometimes used in the literature. In fact, as with some other concepts, it is a comparatively recent departure to offer a definition at all.

If one begins with the premise that mathematical terminology should have a firm foundation in intuition it is perhaps difficult to justify the definition we shall offer, since a Fréchet surface is ultimately defined as a certain class of mappings, and is not, strictly speaking, a geometrical object. Why then preserve the word surface in speaking of the classes we shall study? The reason is principally historical. On the other hand, though it is quite true that the layman would hardly recognize the objects of our attention as surfaces, we propose to show that the definition is a natural consequence of a desire for rigor in the intuitive approach.

2. The geometric concept

There can be no question but that the term surface was first applied to certain point sets. The study of geometry included the study of certain objects known as surfaces. It was decided on the basis of intuition which objects were surfaces and which were not. In a sense we have a catalogue of surfaces in the early history of the subject.

It became generally accepted that a point set was a surface if it was in some sense *flat* at every point. These intuitive concepts were made rigorous with the advent of topology and the notion of two dimensional manifolds.

So much for the geometric and terminological implications of the word. Around the time of Descartes, matters in a sense stood at the place where the body of available mathematicians passed judgment on each point set and pro-

[1] Presented to the American Mathematical Society in April of 1940, 1942 and 1943 under the title "On parametric representations of surfaces I, II and III." The term *Fréchet surface* is due to a suggestion of Radó.

nounced it guilty or innocent of being a surface. It cannot be asserted that the introduction of analytic methods was an unadulterated blessing.

3. The analytic approach

An equation of the type $x^2 + y^2 + z^2 = 1$ determines a point set which was universally accepted as a surface. The equation is a means of *defining* the surface. From this there was probably a gradual transition to speaking of the surface $x^2 + y^2 + z^2 = 1$. Gradually that which was employed as a description of an object came to replace the object itself.

This alone led to no serious confusion, but there came a day in which one began to speak of two equations, which to all appearances were entirely *dissimilar*, as defining the *same* surface. It would be an interesting but no doubt impossible piece of historical research to discover at just what juncture this occurred. Certainly the idea was well entrenched by the time Darboux wrote his treatise on surfaces. There he speaks of a surface being defined in two ways—parametrically and non-parametrically.

Suppose one were given two sets of continuous functions

$$\text{(a)}\quad \begin{aligned} x &= x(u, v) \\ y &= y(u, v) \\ z &= z(u, v) \end{aligned} \quad \text{and} \quad \text{(b)}\quad \begin{aligned} x &= \xi(u, v) \\ y &= \eta(u, v) \\ z &= \zeta(u, v) \end{aligned}$$

where in each case the point (u, v) is in the unit square $X: 0 \leqq u \leqq 1, 0 \leqq v \leqq 1$.

The following array of questions presented itself. I. Does the triple of functions (a) define a surface, or, what was presumably the same question, is the triple a surface? II. Is the triple (b) a surface? III. If both questions are answered in the affirmative, are the surfaces the same?

To decide the matter, a student would have to consult the point sets determined by the triples and decide by consulting some catalogue whether these were surfaces or not. Then by some standard he would have to judge whether the surfaces were the same. The standard of judgment was probably never so lax as to allow one to say that the surfaces were the same merely if the point sets were identical. The surfaces were usually called the same if one could obtain the second triple from the first by suitable *change of parameter*.

Faced with the decisions to be made in this matter it became the practice to eliminate a consideration of the point set—and so, apparently, to discard geometry and topology—and state that two triples did or did not define the same surface purely on the basis of a change of parameter.

4. The notion of equivalence

All that remains to be done is to trace the growth of rigor in the concept of two triples *defining* the same surface.

For the sake of future convenience let us think of the triples (a) and (b) as mappings from X onto some point set Y in E_3. The points of X are designated by small Latin letters.

Whatever is to be meant by the statement that two mappings define the same surface several requirements should be satisfied. The mapping (a) should define the same surface as the mapping (a). If the mapping (a) defines the same surface as the mapping (b) then the mapping (b) should define the same surface as (a). Finally if three mappings are so related that the first defines the same surface as the second, and the second as the third, then the first and third should define the same surface. In other words the relation "*defines the same surface as*" is reflexive, symmetric and transitive; it is an *equivalence* relation over the class of continuous mappings.

In a very real sense it may be said that this paper is the study of such equivalence relations.

There is nothing new about this point of view and comparable statements have appeared in the literature, though only in recent years. On the other hand there is nothing extraordinary about this since the general use of equivalence relations essentially had to await the development of modern algebra.

Radó has recently called my attention to a definition of curves in the spirit of this section. It occurs in a paper by Kneser in the Mathematische Zeitschrift for 1926, volume 25, page 363.

In this connection it can be argued that Lebesgue used an equivalence relation which we define below though the terminology of equivalence is certainly not his.

The mapping $f_1(X)$ is *Lebesgue equivalent* to the mapping $f_2(X)$ if and only if there is a homeomorphism $h(X) = X$ such that $f_1(x) = f_2(h(x))$ for $x \epsilon X$.

There are other standards by which one can decide if two maps "define the same surface" but before going into them it is perhaps well to note the limitations imposed upon the phrase in quotation marks. It certainly cannot be interpreted to mean that the concept of surface has been defined.

5. Surface and representation

The next step is inevitable. The collection of maps which "define the same surface" is an *equivalence class*; this equivalence class is defined to be a *surface*.

Explicitly the situation is this. Suppose that we are given an equivalence relation $\sim$ over the class $\mathcal{F}$ of mappings $f(X)$. This equivalence partitions $\mathcal{F}$ into mutually exclusive equivalence classes $[f]$. Each class may be considered as having been generated as follows. If f_1 is a map then the collection of all maps f such that $f \sim f_1$ is the equivalence class $[f_1]$.

It follows that $f_1 \epsilon [f_1]$ since the relation is reflexive; and $f_1 \sim f_2$ implies that $[f_1] = [f_2]$ since the relation is symmetric and transitive.

Hence a surface is a concept relative to some equivalence relation. A surface relative to the equivalence relation $\sim$ is an equivalence class $[f]$. If we use Lebesgue equivalence we obtain Lebesgue surfaces, if we use an equivalence due to Fréchet we obtain Fréchet surfaces, if we use a Blank equivalence we obtain Blank surfaces.

In the future we shall not say that two mappings define the same surface but that they *generate* or *represent* the same surface if the mappings are in the same

equivalence class. In line with this convention a *representation* of a surface $[f]$ is a member of the equivalence class $[f]$.

It will have been noticed that the set X, or *range*, is a fixed set—the unit square. This has been selected in connection with the triples (a) and (b) of §3 for analytic convenience. Generalizations in two directions are possible and are in the literature. The range X can be taken to be the closure of a region in the plane whose boundary is a finite number of Jordan curves. Or the range X can be a 2-sphere. In the event the range X were an arc, however, the equivalence classes would *not* be called surfaces, but *curves*. In the event the range X were a cube, the equivalence classes would be called *solids*.

To make the matter explicit, the name of the equivalence class depends upon the topological character of the range. Thus it is customary to speak of surfaces of the topological type of a 2-sphere, meaning that the range is a 2-sphere, or surfaces of the topological type of a torus, meaning that the range is a torus.

Another generalization should be mentioned. The statement that $[f]$ is a Blank surface of the topological type of the range X means that all the mappings in the Blank equivalence class $[f]$ are mappings with the fixed range X. The general Blank equivalence, however, will invariably have meaning as long as the mappings are from ranges X_1 and X_2 of the same topological type. For example, it is possible to say that the mapping $f_1(X_1)$ is Lebesgue equivalent to the mapping $f_2(X_2)$ if and only if there is a homeomorphism $h(X_1) = X_2$ such that $f_1(x^1) = f_2(h(x^1))$, for $x^1 \epsilon X_1$.

6. Fréchet surface

This paper is mainly concerned with Fréchet surfaces. In other words its principal object is the examination of Fréchet equivalence which we now propose to define.

The mapping $f_1(X_1)$ is Fréchet equivalent to the mapping $f_2(X_2)$ if and only if for every positive ϵ there is a homeomorphism $h_\epsilon(X_1) = X_2$ such that $\rho\{f_1(x^1), f_2(h_\epsilon(x^1))\} < \epsilon$ for $x^1 \epsilon X_1$.

The notion of Fréchet equivalence is not as simple as the notion of Lebesgue equivalence. The latter may be thought of as an exact matching of the maps by means of a homeomorphism while the former is a matching within a controlled error of ϵ.

It is not difficult to construct examples to show that two mappings which are Fréchet equivalent are not, in general, Lebesgue equivalent. On the other hand it is quite obvious that Lebesgue equivalence implies Fréchet equivalence.

7. The main problem

The collection of maps which are Lebesgue equivalent to a given mapping $f(X_1)$ is easy to visualize. It is simply the collection of mappings $g(X_2) \equiv f(h(X_2))$ where $h(X_2) = X_1$ is a homeomorphism from X_2 onto X_1. The collection of mappings which are Fréchet equivalent to $f(X_1)$ contains all those given

by the formula above, and others. It is not so easy to see how the others are constructed.

This brings us to what is perhaps the fundamental problem of the paper. *Given one representation f of a Fréchet surface $[f]$ find all the other representations.*

7a. An alternate point of view

This introduction would not be complete without the addition of some comments regarding a different approach to the concept of surface. These remarks are the consequence of a conversation with Radó.

In precisely the same fashion as a physicist uses mathematical models to correspond to, but not replace, physical phenomena, a mathematician constructs mathematical models corresponding to mathematical phenomena. For example, the real number system has been used for centuries in an intuitive fashion and is so used today by many people who rely on mathematics. Thus the real number system as used for centuries is a mathematical phenomenon for which only recently mathematicians have constructed mathematical models. A second example is the intuitive concept of connectedness, a mathematical phenomenon for which Lennes constructed a mathematical model. In all such cases the mathematical model is abstract and is to serve the expert, not the layman.

Similarly, it can be argued, the words curve and surface were used intuitively; in geometry applied to figures, in analysis, to equations. A mathematical model for the geometric type of usage is found in the concept of one and two dimensional manifolds. A mathematical model for the second type of usage is the concept of Fréchet curves and surfaces as employed in this paper.

Chapter II

Topological Preliminaries

This chapter is concerned with a short statement of some of the topological concepts required in the body of the paper. In a general way we will assume a familiarity with Whyburn [10] for the point set theoretic portions, and Alexandroff-Hopf [2] for the combinatorial theory.

8. Point set topology

The list of theorems is not intended to be exhaustive and to save space only the key propositions are mentioned.

All spaces will be metric and any mapping is understood to be continuous.

Theorem 8.1. *If $f(X) = Y$ is a biunique mapping and X is compact then it is a homeomorphism.*

Theorem 8.2. (Eilenberg-Whyburn Factor Theorem). *If X is compact then any mapping $f(X) = Y$ can be factored,*

$$f(X) = lm(X)$$

where $m(X) = \Sigma$ is monotone and $l(\Sigma) = Y$ is light.

THEOREM 8.3. *In order that the mapping $l(X) = Y$ be light it is necessary and sufficient that for every $\epsilon > 0$ there is a $\delta > 0$ such that if C is a continuum of diameter $< \delta$ in Y then any component of $f^{-1}(C)$ has a diameter $< \epsilon$.*

In connection with Peano space theory, where Kuratowski-Whyburn [11] is sufficient for our purposes, we need the following facts.

THEOREM 8.4. *The true cyclic elements of a Peano space form a null sequence.*

THEOREM 8.5. *If A is an A-set in a Peano space X then the components of $X - A$ form a null sequence and each has a single point for frontier.*

Suppose that A is an A-set, and consider the mapping

$$\mu(A; x) = \begin{cases} x \text{ if } x \in A \\ F(G) \text{ if } x \in G \text{ a component of } X - A \end{cases} \qquad \text{(II, 1)}$$

This mapping is a monotone retraction of X onto A and is of fundamental importance.

THEOREM 8.6. *For the Peano space X there is a sequence of cyclic chains $C(p_n, q_n)$ such that*

1) $\bigcup_1^n C(p_i, q_i) \equiv A_n$ *is an A-set for each n.*
2) $A_n \cdot C(p_{n+1}, q_{n+1}) = q_{n+1}$ *for each n.*
3) $\bigcup A_n \equiv H$ *is dense in X*
4) *If d_n is the supremum of the diameters of the components of $X - A_n$, then $d_n \to 0$.*
5) $d\{C(p_n, q_n)\} \to 0$.

We are primarily interested in special Peano spaces—notably 2-spheres. It is convenient to have a topological characterization of the 2-sphere and we mention a beautiful theorem of Kuratowski: *a 2-sphere is a non-degenerate cyclic Peano space whose non-cutting continua are unicoherent. By means of this and standard properties of monotone maps it is easily shown that:*

THEOREM 8.7. *The monotone image of a 2-sphere is a cactoid; i.e., a Peano space whose true cyclic elements are 2-spheres.*

THEOREM 8.8. *Any cactoid is the monotone image of a 2-sphere.*

Nothing has been done, apparently, with upper semi-continuous collections in this discussion, but their usefulness has been implied everywhere. We will encounter them in the future. Before concluding this paragraph we should mention an often applied theorem of Janisewski.

THEOREM 8.9. *If $\{C_n\}$ is a sequence of continua in a Peano space X and* lim inf $C_n \neq 0$, *then* lim sup C_n *is a continuum.*

9. Combinatorial topology

The principal concept here is that of orientation and degree of a mapping (Alexandroff-Hopf [2]).

We shall be interested mainly in the degree of a mapping from an oriented 2-sphere onto another. If f is the mapping the notation employed is Dgr f.

THEOREM 9.1. *If X_1 and X_2 are oriented 2-spheres, $f_n(X_1) = X_2$ while $f_n(X_1)$ converges uniformly to $f(X_1)$, then for sufficiently large n,* Dgr f_n = Dgr f.

THEOREM 9.2. *If $h(X_1) = X_2$ is a homeomorphism of an oriented 2-sphere into another, then* Dgr $h = \pm 1$.

If J_i is a (closed) 2-cell on the oriented 2-sphere X_i, $i = 1, 2$, and we have a homeomorphism $h(J_1) = J_2$, then we can speak of the degree of h in terms of the orientations on X_1 and X_2 since these orientations induce orientations on J_1 and J_2.

THEOREM 9.3. *Suppose that $J_{i1}, \cdots, J_{in}$ are 2-cells on an oriented 2-sphere X_i, and $J_{ik}\cdot J_{il} = 0$, $k \neq l$, $i = 1, 2$. If there is a homeomorphism $f_k(J_{1k}) = J_{2k}$ of degree 1 (-1) for $k = 1, \cdots, n$, then there is a homeomorphism $f(X_1) = X_2$ which agrees with f_k on J_{1k} and has degree 1 (-1).*

This theorem is certainly well known though apparently not stated explicitly in the literature. It can be proved as a consequence of some results of Adkisson and MacLane [1] on extensions of homeomorphisms. Some remarks will be made to indicate the connection.

There are $(n - 1)$ arcs $\alpha_1, \cdots, \alpha_{n-1}$ such that

1) For each k the arc α_k has the end point a_k on J_{1k} and the other end point b_k on J_{1k+1}, but no other point on any of the J_1's.

2) $\alpha_k\cdot\alpha_l = 0, \quad k \neq l.$

Let $f_k(a_k) = p_k$ and $f_{k+1}(b_k) = q_k$. Then there is a collection of arcs $\beta_1, \cdots, \beta_{n-1}$ such that the β's have the same properties on X_2 as the α's on X_1.

Let P_i be the set of points in the 2-cells and arcs of X_i, $i = 1, 2$. The sets P_1 and P_2 are Peano spaces on the oriented 2-spheres X_1 and X_2.

Let $g(\alpha_k) = \beta_k$ be a homeomorphism such that $g(a_k) = p_k$, $g(b_k) = q_k$, $k = 1, \cdots, (n - 1)$. Let $h(P_1) = P_2$ be a mapping which agrees with g on α_k and f_l on J_{1l}, for every possible k and l.

The mapping h is a homeomorphism which satisfies the triod condition of Adkisson and MacLane and the theorem follows.

CHAPTER III

MONOTONE MAPS AND HOMEOMORPHISMS

In this chapter it will be our object to make a preliminary application of some of the topological tools mentioned.

The final objective, roughly speaking, is to approximate certain monotone transformations uniformly by means of homeomorphisms. The monotone transformations in question are from one 2-sphere onto another. The degree of such mappings can then be calculated at once by 9.1 and 9.2.

10. A modification theorem

THEOREM 10.1. *If $m(X) = Y$ is a monotone mapping from the 2-sphere X onto the 2-sphere Y, and G is the interior of a 2-cell on Y while $G^* = m^{-1}(G)$, then there is a monotone mapping $h(X) = Y$ which is a homeomorphism on G^* and elsewhere agrees with m.*

PROOF. Consider a decomposition Σ of X consisting of the sets $m^{-1}(y)$ for

$y \in Y - G$ and single points elsewhere. This decomposition is upper semi-continuous and so there is a unique topologization of Σ compatible with the condition that the related transformation $\phi(X) = \Sigma$ be continuous. We recall that the transformation ϕ maps each point of X onto the element of Σ which, as a point set in X, contains it. Since each set of the decomposition is a continuum the mapping ϕ is monotone. Moreover the decomposition is free of sets which cut X and so the cactoid Σ is a single 2-sphere.

Define a mapping from $Y - G$ into Σ by the following formula:

$$\theta(y) = \phi m^{-1}(y), \qquad y \in Y - G.$$

Standard results indicate that this mapping is a homeomorphism, and so $\theta(Y - G)$ is a 2-cell on the 2-sphere Σ. By a theorem of Schoenfliess the homeomorphism can be extended to a homeomorphism of Y onto Σ.

Let $\theta(Y) = \Sigma$ represent the extended mapping.

Now consider $h(X) = \theta^{-1}\phi(X)$. We notice that

1) h is a homeomorphism on G^*.

$$2)\ \left.\begin{aligned} h(x) &= \theta^{-1}\phi(x) \\ &= m\phi^{-1}\phi(x) \\ &= m(x) \end{aligned}\right\} x \in X - G^*.$$

11. Approximation theorems

The theorem just proved is fundamental and several applications are made. Perhaps the easiest to state is the following.

Approximation Theorem 11.1. *If $m(X) = Y$ is a monotone map from a 2-sphere X onto a 2-sphere Y and $\epsilon > 0$, then there is a homeomorphism $h(X) = Y$ such that $\rho\{h(x), m(x)\} < \epsilon$ for $x \in X$.*

Proof. First suppose that X and Y are geometrical spheres rather than being merely topological. A triangulation T of X is a decomposition of X into spherical triangles $\Delta_1, \cdots, \Delta_n$ such that: 1) the sides of each triangle are arcs of great circles, 2) two distinct triangles which have a non-vacuous intersection meet in a single common side, or, failing this, in a single common vertex. Let $|T| = \max d(\Delta_n)$. An η-triangulation T is a triangulation with the property that $|T| < \eta$.

Suppose that $\epsilon = 4\eta > 0$ and T is an η-triangulation of Y. If the triangles are $\Delta_1, \cdots, \Delta_n$ then use the notation $T_1, \cdots, T_n$ to denote their interiors. In each triangle select an interior point and do the same for each side. (In the last case the point will be interior to the side relative to the side). Suppose that the points in the sides are labelled $y_1, \cdots, y_r$. Draw three arcs in each Δ; namely, from the interior point of the triangle to the interior points of the three sides. The sphere Y is partitioned into Jordan regions $R_1, \cdots, R_m$ each bounded by a finite number of these arcs. This is the dual decomposition. Each region contains exactly one vertex of the triangulation. The diameter of the largest of these regions is 2η.

The plan of attack is immediate. We apply Theorem 10.1 to obtain a mapping

$h_1(X) = Y$ which agrees with m except in $m^{-1}(T_1) = T_1^*$ where it is a homeomorphism $h_1(T_1^*) = T_1$. Now apply the theorem to $h_1(X) = Y$ to obtain a mapping $h_2(X) = Y$ which agrees with h_1 except in $h_1^{-1}(T_2) = T_2^*$ where it is a homeomorphism $h_2(T_2^*) = T_2$. Proceeding in this fashion we arrive at $h_n(X) = Y$ which agrees with h_{n-1} except in $h_{n-1}^{-1}(T_n) = T_n^*$ where it is a homeomorphism $h_n(T_n^*) = T_n$.

It follows that $h_n(X) = Y$ agrees with $m(X) = Y$ on $m^{-1}(F(T_1 + \cdots + T_n))$ and $h_n(T_k^*) = T_k$ is a homeomorphism for each $k = i, \cdots, n$. This implies $\rho\{h(x), m(x)\} < \eta \ x \in X$.

The method of changing m with respect to the triangles of the triangulation can now be applied to h_n working through the Jordan regions of the dual decomposition. This will yield a map g which agrees with h_n on $h_n^{-1}(F(R_1 + \cdots + R_m))$ and $g(h_n^{-1}(R_j)) = R_j$ is a homeomorphism for each $j = 1, \cdots, m$. Moreover, $\rho\{h_n(x), g(x)\} < 2\eta$ and so $\rho\{g(x), m(x)\} < 3\eta$ for $x \in X$.

The mapping g will fail to be a homeomorphism, if at all, only by virtue of the fact that it agrees with m on the set $m^{-1}(y_1 + \cdots + y_r)$.

For each q select a Jordan region G_q such that $y_q \in G_q \ \bar{G}_j \cdot \bar{G}_k = 0$ for $j \neq k$, and $\max d(G_q) < \eta$. Repeat the twice used process on the collection $G_1, \cdots, G_r$. The result is a homeomorphism $h(X) = Y$ with the property that $\rho\{h(x), m(x)\} < 4\eta = \epsilon$ for $x \in X$.

The case for the general 2-sphere can easily be reduced to the argument presented above.

Remark. A similar theorem for 2-cells is stated by Morrey [6, p. 39]. Morrey's proof is open to objection, and since the proof above uses the characterization of a 2-sphere, the theorem itself is perhaps deeper than the elementary comments of Morrey might lead one to believe.

As an immediate consequence we have the useful

COROLLARY 11.2. *If X and Y are oriented, then* Dgr $m = \pm 1$.

Actually the theorem is not quite in the best form for the final applications and we now state an easily proved modification.

THEOREM 11.3. *If $m(X) = Y$ is a monotone map from a 2-sphere X onto a 2-sphere Y, $p_1, \cdots, p_n$ is any finite set of points on Y and $\epsilon > 0$, then there is a map $\mu(X) = Y$ which agrees with m on $m^{-1}(p_1 + \cdots + p_n)$, is a homeomorphism elsewhere, and $\rho\{m(x), \mu(x)\} < \epsilon$ for $x \in X$.*

12. The matching theorem

The theorem of this paragraph can reasonably be described as one of the key theorems of the paper. It concerns clusters.

A cactoid Σ is a *cluster* if $\Sigma = S_1 + \cdots + S_n$ where each S_k is a 2-sphere. Suppose that the collection of cut points to be $E = p_1 \cdots p_m$. Each will be common to two or more true cyclic elements.

Suppose we have a monotone map $m(X) = \Sigma$ from an oriented 2-sphere X onto Σ. If S is a true cyclic element of Σ let $m(S; X)$ be $m(X)$ followed by the

monotone retraction of X onto S, (II, 1). This product map is a monotone map from X onto S and if S is oriented it has degree 1 or -1, (11.2).

MATCHING THEOREM 12.1. *If for* $i = 1, 2$,

1) $m_i(X_i) = \Sigma$ *is a monotone map.*

2) X_i *is an oriented 2-sphere.*

3) Σ *is a cluster.*

4) *The 2-spheres of* Σ *are so oriented that* Dgr $m_1(S; X_1) = 1\ (-1)$ *for each* S *of* Σ.

5) *With the orientation on* S *determined by condition* 4) *we have* Dgr $m_2(S; X_2) = 1$, *for each* S *of* Σ.

6) $\epsilon > 0$.

Then there is a homeomorphism $h_\epsilon(X_1) = X_2$ *such that*

1) $\rho\{m_1(x^1),\ m_2(h_\epsilon(x^1))\} < \epsilon,\ x' \in X_1$.

2) Dgr $h_\epsilon = 1\ (-1)$.

Note that if -1 is read in part 4 of the hypothesis, then -1 is read in part 2 of the conclusion. This abbreviation is often used.

PROOF. Let $S_1, \cdots, S_n$ be the true cyclic elements and $p_1, \cdots, p_m$ the cut points of Σ. Suppose that the minimum distance between cut points of the cluster is $>4\eta = \epsilon$, and, moreover, the diameter of the smallest S_k is $>4\eta$. On each 2-sphere S_k there will be at least one point of E unless there is but a single S_k and then the theorem is too simple to consider in detail as a special case.

Select a point of E on S_k. There is a Jordan region on S_k of diameter $<\eta$ which contains this point. Such regions are chosen for each point of E on S_k, $k = 1, \cdots, n$.

The situation is now this: Each point of E belongs to several 2-spheres of Σ and in each 2-sphere of which it is an element it lies in a Jordan region. If the regions are $D_1, \cdots, D_s$ then the statement $\bar{D}_j \cdot \bar{D}_k \neq 0$, $j \neq k$ implies that $\bar{D}_j \cdot \bar{D}_k = p$, some point of E. The sum of all the D's containing p is a neighborhood of p in Σ.

Suppose the boundary of D_k is J_k ($k = 1, \cdots, s$), the boundary being taken relative to the 2-sphere in which D_k lies.

Consider the set $\Sigma - \bigcup_1^s J_k$. This is an open set with $m + n$ components. Each true cyclic element contains exactly one of these components, each cut point is contained in one of them.

Suppose R_l is the component containing p_l. It consists of the sum of as many D's as there are 2-spheres having p in common. Let S_k^* be the component contained in S_k. This is bounded by as many Jordan curves as there are points of E contained in S_k.

Let us now return to the mapping $m_i(X_i) = \Sigma$. For each i, let the mapping $m_i(S_k; X_i) = S_k$ be denoted by $m_{ik}(X_i) = S_k$. Using 11.3 we obtain a mapping $\mu_{ik}(X_i) = S_k$ such that:

1) μ_{ik} agrees with m_i on $m_i^{-1}(S_k, E)$.

2) Is a homeomorphism elsewhere.

3) Dgr μ_{ik} = Dgr m_{ik}.

4) $\rho\{\mu_{ik}(x^1), m_{ik}(x^1)\} > \eta$ for $x^1 \epsilon m_i^{-1}(S_k - E)$.
The set $m_i^{-1}(S_k - E)$ will be denoted by x_{ik}.
For each i define

$$\mu_i(x^1) = \begin{cases} \mu_{ik}(x^i) & \text{if} \quad x^i \epsilon X_{ik} \\ m_i(x^i) & \text{if} \quad x^i \epsilon m_i^{-1}(E) \end{cases} \qquad \text{(III, 1)}$$

We now have a means of obtaining a homeomorphism from $\mu_1^{-1}(S_k^*)$ onto $\mu_2^{-1}(S_k^*)$, in fact, define

$$h(x^1) = \mu_2^{-1}\mu_1(x^1) \quad \text{for} \quad x^1 \epsilon \mu_1^{-1}(\bar{S}_k^*), \qquad k = 1, \cdots, n.$$

Then Dgr $h = 1\ (-1)$.

This mapping is not defined over all of X_1 nor is the image all of X_2. The telling point is that it can be extended to a homeomorphism from X_1 onto X_2. This is the important point, the extension need fulfill no additional properties.

We shall describe the extension from the set $\mu_1^{-1}(\bar{R}_l)$ onto $\mu_2^{-1}(\bar{R}_l)$. With no loss of generality we may suppose that R_l consists of Jordan regions on r 2-spheres $S_1, \cdots, S_r$. Then $F(R_l) = J_1 + \cdots + J_r$ where the J's are Jordan curves, one on each of the 2-spheres.

For each i the region $\mu_i^{-1}(R_l)$ will have r Jordan curves on its boundary, $C_{i1}, \cdots, C_{ir}$. (Suppose that the corresponding Jordan regions complementary to $\mu_i^{-1}(R_l)$ are $G_{i1}, \cdots, G_{ir}$.)

Consider the mapping $\mu_{ik}(X_i) = S_k$ for each k. The points of E on S_k are, let us say, $p_k = q_1, \cdots, q_t$. On S_k each of these is in a Jordan region already chosen at the beginning of the argument, one point to a region. (These regions are the D's).

Alter μ_{ik} by the technique of 10.1 with respect to these D's to obtain a homeomorphism $\phi_{ik}(X_i) = S_k$. We need only that Dgr ϕ_{ik} = Dgr μ_{ik}. Define

$$f_k(x^1) = \phi_{2k}^{-1}\phi_{1k}(x^1).$$

This is a homeomorphism, and

$$\begin{aligned} \text{Dgr}\, f_k &= \text{Dgr}\, \phi_{2k}^{-1} \cdot \text{Dgr}\, \phi_{1k} = \text{Dgr}\, \phi_{2k} \cdot \text{Dgr}\, \phi_{1k} \\ &= \text{Dgr}\, \mu_{2k} \cdot \text{Dgr}\, \mu_{1k} = \text{Dgr}\, m_{2k} \cdot \text{Dgr}\, m_{1k} \\ &= 1\ (-1). \end{aligned}$$

We notice that f_k agrees with h on C_{1k}.

The f's play the rôle of a tool. We have 2-cells $\bar{G}_{11}, \cdots, \bar{G}_{1k}$ taken by $f_1, \cdots, f_r$ onto $\bar{G}_{21}, \cdots, \bar{G}_{2r}$ respectively. These 2-cells are disjoint and the degree of each f_k is $1\ (-1)$. By 9.3 there is a homeomorphism $g_l(X_1) = X_2$ such that $g_l(\bar{G}_{1k}) = f_k(\bar{G}_{1k})$, $k = 1, \cdots, r$, and Dgr $g_l = 1\ (-1)$. We shall be interested in $g_l(X_1) = X_2$ only on the set $\overline{\mu_1^{-1}(R_l)} = \mu_1^{-1}(\bar{R}_l)$. The letters g_l now refer *only* to this partial mapping. The maps h and g_l agree on the only set where they are both defined, the collection of Jordan curves $C_{11}, \cdots, C_{1r}$. If this is done for each $l = 1, \cdots, m$ then the composite mapping $h, g_1, \cdots, g_m$ is a homeomorphism of degree $1\ (-1)$ which we will name $h_\epsilon(X_1) = X_2$.

monotone retraction of X onto S, (II, 1). This product map is a monotone map from X onto S and if S is oriented it has degree 1 or -1, (11.2).

MATCHING THEOREM 12.1. *If for* $i = 1, 2$,

1) $m_i(X_i) = \Sigma$ *is a monotone map.*

2) X_i *is an oriented 2-sphere.*

3) Σ *is a cluster.*

4) *The 2-spheres of* Σ *are so oriented that* Dgr $m_1(S; X_1) = 1\ (-1)$ *for each* S *of* Σ.

5) *With the orientation on* S *determined by condition* 4) *we have* Dgr $m_2(S; X_2) = 1$, *for each* S *of* Σ.

6) $\epsilon > 0$.

Then there is a homeomorphism $h_\epsilon(X_1) = X_2$ *such that*

1) $\rho\{m_1(x^1), m_2(h_\epsilon(x^1))\} < \epsilon$, $x' \epsilon X_1$.

2) Dgr $h_\epsilon = 1\ (-1)$.

Note that if -1 is read in part 4 of the hypothesis, then -1 is read in part 2 of the conclusion. This abbreviation is often used.

PROOF. Let $S_1, \cdots, S_n$ be the true cyclic elements and $p_1, \cdots, p_m$ the cut points of Σ. Suppose that the minimum distance between cut points of the cluster is $>4\eta = \epsilon$, and, moreover, the diameter of the smallest S_k is $>4\eta$. On each 2-sphere S_k there will be at least one point of E unless there is but a single S_k and then the theorem is too simple to consider in detail as a special case.

Select a point of E on S_k. There is a Jordan region on S_k of diameter $<\eta$ which contains this point. Such regions are chosen for each point of E on S_k, $k = 1, \cdots, n$.

The situation is now this: Each point of E belongs to several 2-spheres of Σ and in each 2-sphere of which it is an element it lies in a Jordan region. If the regions are $D_1, \cdots, D_s$ then the statement $\bar{D}_j \cdot \bar{D}_k \neq 0$, $j \neq k$ implies that $\bar{D}_j \cdot \bar{D}_k = p$, some point of E. The sum of all the D's containing p is a neighborhood of p in Σ.

Suppose the boundary of D_k is J_k $(k = 1, \cdots, s)$, the boundary being taken relative to the 2-sphere in which D_k lies.

Consider the set $\Sigma - \bigcup_1^s J_k$. This is an open set with $m + n$ components. Each true cyclic element contains exactly one of these components, each cut point is contained in one of them.

Suppose R_l is the component containing p_l. It consists of the sum of as many D's as there are 2-spheres having p in common. Let S_k^* be the component contained in S_k. This is bounded by as many Jordan curves as there are points of E contained in S_k.

Let us now return to the mapping $m_i(X_i) = \Sigma$. For each i, let the mapping $m_i(S_k; X_i) = S_k$ be denoted by $m_{ik}(X_i) = S_k$. Using 11.3 we obtain a mapping $\mu_{ik}(X_i) = S_k$ such that:

1) μ_{ik} agrees with m_i on $m_i^{-1}(S_k, E)$.

2) Is a homeomorphism elsewhere.

3) Dgr μ_{ik} = Dgr m_{ik}.

4) $\rho\{\mu_{ik}(x^1), m_{ik}(x^1)\} > \eta$ for $x^1 \epsilon m_i^{-1}(S_k - E)$.
The set $m_i^{-1}(S_k - E)$ will be denoted by x_{ik}.
For each i define

$$\mu_i(x^1) = \begin{cases} \mu_{ik}(x^i) & \text{if} \quad x^i \epsilon X_{ik} \\ m_i(x^i) & \text{if} \quad x^i \epsilon m_i^{-1}(E) \end{cases} \qquad \text{(III, 1)}$$

We now have a means of obtaining a homeomorphism from $\mu_1^{-1}(S_k^*)$ onto $\mu_2^{-1}(S_k^*)$, in fact, define

$$h(x^1) = \mu_2^{-1}\mu_1(x^1) \quad \text{for} \quad x^1 \epsilon \mu_1^{-1}(\bar{S}_k^*), \qquad k = 1, \cdots, n.$$

Then Dgr $h = 1$ (-1).

This mapping is not defined over all of X_1 nor is the image all of X_2. The telling point is that it can be extended to a homeomorphism from X_1 onto X_2. This is the important point, the extension need fulfill no additional properties.

We shall describe the extension from the set $\mu_1^{-1}(\bar{R}_l)$ onto $\mu_2^{-1}(\bar{R}_l)$. With no loss of generality we may suppose that R_l consists of Jordan regions on r 2-spheres $S_1, \cdots, S_r$. Then $F(R_l) = J_1 + \cdots + J_r$ where the J's are Jordan curves, one on each of the 2-spheres.

For each i the region $\mu_i^{-1}(R_l)$ will have r Jordan curves on its boundary, $C_{i1}, \cdots, C_{ir}$. (Suppose that the corresponding Jordan regions complementary to $\mu_i^{-1}(R_l)$ are $G_{i1}, \cdots, G_{ir}$.)

Consider the mapping $\mu_{ik}(X_i) = S_k$ for each k. The points of E on S_k are, let us say, $p_k = q_1, \cdots, q_t$. On S_k each of these is in a Jordan region already chosen at the beginning of the argument, one point to a region. (These regions are the D's).

Alter μ_{ik} by the technique of 10.1 with respect to these D's to obtain a homeomorphism $\phi_{ik}(X_i) = S_k$. We need only that Dgr ϕ_{ik} = Dgr μ_{ik}. Define

$$f_k(x^1) = \phi_{2k}^{-1}\phi_{1k}(x^1).$$

This is a homeomorphism, and

$$\begin{aligned} \text{Dgr } f_k &= \text{Dgr } \phi_{2k}^{-1} \cdot \text{Dgr } \phi_{1k} = \text{Dgr } \phi_{2k} \cdot \text{Dgr } \phi_{1k} \\ &= \text{Dgr } \mu_{2k} \cdot \text{Dgr } \mu_{1k} = \text{Dgr } m_{2k} \cdot \text{Dgr } m_{1k} \\ &= 1\ (-1). \end{aligned}$$

We notice that f_k agrees with h on C_{1k}.

The f's play the rôle of a tool. We have 2-cells $\bar{G}_{11}, \cdots, \bar{G}_{1k}$ taken by $f_1, \cdots, f_r$ onto $\bar{G}_{21}, \cdots, \bar{G}_{2r}$ respectively. These 2-cells are disjoint and the degree of each f_k is 1 (-1). By 9.3 there is a homeomorphism $g_l(X_1) = X_2$ such that $g_l(\bar{G}_{1k}) = f_k(\bar{G}_{1k})$, $k = 1, \cdots, r$, and Dgr $g_l = 1$ (-1). We shall be interested in $g_l(X_1) = X_2$ only on the set $\overline{\mu_1^{-1}(R_l)} = \mu_1^{-1}(\bar{R}_l)$. The letters g_l now refer *only* to this partial mapping. The maps h and g_l agree on the only set where they are both defined, the collection of Jordan curves $C_{11}, \cdots, C_{1r}$. If this is done for each $l = 1, \cdots, m$ then the composite mapping $h, g_1, \cdots, g_m$ is a homeomorphism of degree 1 (-1) which we will name $h_\epsilon(X_1) = X_2$.

If $x^1 \epsilon \mu_1^{-1}(\bar{S}_k^*)$ then $\mu_2(h_\epsilon(x^1)) = \mu_2\mu_2^{-1}\mu_1(x^1) = \mu_1(x^1)$. Hence $\rho\{\mu_1(x^1), \mu_2 h_\epsilon(x^1)\} = 0$. This is true for $k = 1, \cdots, n$.

If $x^1 \epsilon \mu_1^{-1}(R_l)$ then $h_\epsilon(x^1) = g_l(x^1) \epsilon \mu_2^{-1}(R_l)$ and hence $\mu_1(x^1)$ and $\mu_2 g_l(x^1)$ are both points in R_l which is of diameter $< 2\eta$.

Hence $\rho\{\mu_1(x^1), \mu_2 h_\epsilon(x^1)\} < 2\eta$.

Combining these facts we have $\rho\{\mu_1(x^1), \mu_2 h_\epsilon(x^1) < 2\eta$ for $x^1 \epsilon X_1$.

On the other hand $\rho\{\mu_i(x^i), m_i(x^i)\} < \eta, x^i \epsilon X_i, i = 1, 2$.

Therefore $\rho\{m_1(x^1), m_2 h_\epsilon(x^1)\} < 4\eta < \epsilon$ and the proof is complete.

Chapter IV

The Equivalences

In Chapter I we saw that the concept at the heart of surface theory is that of an equivalence. In our efforts towards solving the fundamental problem mentioned in §7 we shall need several types of equivalence relations not mentioned in the introduction. These will be defined and their properties investigated.

13. Fréchet equivalences

The *ordinary Fréchet equivalence* has already been defined in §6. The definition has meaning as long as the spaces are metric. In the next definition, however, it is understood that X_1 and X_2 are oriented 2-spheres (or 2-cells). For convenience let us understand that the range of a mapping will be an oriented 2-sphere unless the contrary is explicitly stated.

A mapping $f_1(X_1) = Y$ is said to be *positively Fréchet equivalent* to a mapping $f_2(X_2) = Y$ if and only if for every $\epsilon > 0$ there is a homeomorphism $h_\epsilon(X_1) = X_2$ such that: 1) $\rho\{f_1(x^1), f_2(h_\epsilon(x^1))\} < \epsilon$ for every x^1 in X, and 2) Dgr $h_\epsilon = 1$.

A mapping $f_1(X_1) = Y$ is *negatively Fréchet related* to a mapping $f_2(X_2) = Y$ if condition 2) above is modified to read that Dgr $h_\epsilon = -1$.

For convenience we employ the notations $f_1F \sim f_2$, $f_1F^+ \sim f_2$, and $f_1F^- \sim f_2$ respectively.

The last of these relations is not an equivalence, but the first two are. Though the name of Fréchet has been attached to all of these relations it is only the first for which he is directly responsible. The second is properly due to McShane [4] and we write $f_1F^+ \sim f_2$ merely for terminological convenience. The third relation appears to have no importance by itself but has considerable bearing on the problem we have in mind.

Suppose we say that $f_1 \sim f_2$ if $f_1F^+ \sim f_2$ or $f_1F^- \sim f_2$. It can be shown that this is a bona fide equivalence relation and the following theorem holds.

Theorem 13.1. *If $f_1F \sim f_2$ then $f_1 \sim f_2$, and conversely.*

14. A counter-example

The three relations so far introduced are to be compared with three which will be defined in §15 by means of the Eilenberg-Whyburn Factor Theorem (8.2). We digress for a moment to make a historical comment.

If we have two mappings $f_1(X_1) = Y$ and $f_2(X_2) = Y$ then there will be two cactoids Σ_1 and Σ_2 together with two monotone maps m_1 and m_2 and two light transformations l_1 and l_2 held together by the following relations:

$$f_i(X_i) = l_i m_i(X_i), \qquad m_i(X_i) = \Sigma_i, \qquad l_i(\Sigma_i) = Y, \qquad i = 1, 2.$$

The mapping $f_1(X_1) = Y$ is said to be *Kerékjártó equivalent* to the mapping $f_2(X_2) = Y$ if and only if there is a homeomorphism $h(\Sigma_1) = \Sigma_2$ such that $\rho\{l_1(\sigma_1), l_2(h(\sigma_1))\} = 0$ for every point σ_1 of Σ_1. (Notation: $f_1K \sim f_2$.)

This relation could be rephrased so as to read that $f_1K \sim f_2$ if and only if l_1 and l_2 are Lebesgue equivalent. (§4).

The notion is due to Kerékjártó [3]. Sometime later, Morrey [5] used the same notion without being aware of Kerékjártó's contribution in this direction. In some respects it is unfortunate that the definition is recorded here at all as we shall merely make passing use of it. On the other hand its historical interest carries considerable weight and deserves mention.

Kerékjártó proved two theorems. The first that if $f_1F \sim f_2$ then $f_1K \sim f_2$. This theorem is true and his proof, though somewhat sketchy, certainly supplies the skeleton of the argument in the more general Theorem 16.5 proved later in this paper. Morrey does not attempt to investigate this question as it has no bearing, except an asthetic one, on the applications he has in mind. The second theorem of Kerékjártó states that if $f_1K \sim f_2$ then $f_1F \sim f_2$. This theorem is also stated by Morrey in substantially the same form and used extensively by him in some beautiful applications to the problem of area [6]. This particular theorem, however, is false and a simple counter-example was exhibited by Youngs in 1940.

The facts, then, are these. A theorem which is not applied is true, but one with extensive applications is false. It can be stated that the impetus for this paper came from a desire to see what could be saved in this situation.

The characterization of Fréchet equivalence in terms of Kerékjártó equivalence would, if true, be a neat result apart from any applications. This cannot be done, but we propose to characterize it together with the other relations $F^+\sim$ and F^- in terms of concepts akin to Kerékjártó equivalence.

Before defining these relations let us look at a counter-example to the theorem of Kerékjártó and Morrey. The example appears to be in as simple a form as possible.

Consider the mapping $f_1(X) = Y$ which consists of a transformation from the oriented 2-sphere X of radius 1 onto a pair of tangent 2-spheres. It is convenient to consider a great circle on X, to call it the equator and speak of the northern and southern hemispheres. Explicitly, the point x on X with latitude θ is moved directly towards the center of X and the image is at a distance $\sin\theta$ from the center. The closed northern hemisphere is thus mapped onto one of the two tangent spheres arbitrarily called the upper sphere; the closed southern hemisphere is mapped onto the other, called the lower sphere.

The mapping $f_2(X) = Y$ is defined as the mapping $f_1(X) = Y$ followed by a

reflection of the upper tangent sphere on a plane containing the center and point of tangency.

Factor each mapping to get $f_i(X) = l_i m_i(X)$, $m_i(X) = \Sigma_i$, $l_i(\Sigma_i) = Y$, $i = 1, 2$. The abstract spaces Σ_1 and Σ_2 will each consist of two tangent 2-spheres.

Define $h(\sigma_1) = l_2^{-1} l_1(\sigma_1)$, $\sigma_1 \epsilon \Sigma_1$.

We notice that $l_2(h(\sigma_1)) = l_1(\sigma_1)$ and so we need merely to check on the fact that h is a homeomorphism to state that $f_1 K \sim f_2$. This follows at once on observing that both l_1 and l_2 are homeomorphisms since every inverse set is a single point and we are dealing with compact spaces, (8.1).

Now suppose that $f_1 F \sim f_2$. Then for each integer n there is a homeomorphism $h_n(X) = X$ compatible with the requirement that $\rho\{f_1(x), f_2(h_n(x))\} < 1/n$, for every x in X.

We shall speak of the northern (southern) cap of X as the spherical zone of points with latitudes not less than $\pi/6N(S)$. In both zones the mappings f_1 and f_2 are homeomorphisms, thus $h_n(x)$ is uniformly close to $f_2^{-1} f_1(x)$ in both the northern and southern caps. (In fact, an easy calculation indicates that $h_n(x)$ is within $2/n$ of $f_2^{-1} f(x)$ in these caps.) But in any event, the mapping $f_2^{-1} f_1(x)$ is a reflection on a plane through the poles of X in the northern cap and hence its degree is -1. On the southern cap $f_2^{-1} f_1(x)$ is the identity and has degree 1.

It follows by Theorems 9.1 and 9.2 that if $f_1 F \sim f_2$ there is a homeomorphism $h(X) = X$ which simultaneously has degree 1 and -1. This is impossible.

15. The modified Kerékjártó equivalences

The ordinary Kerékjártó equivalence fails, as the counter-example indicates, due to the fact that it ignores the matter of orientation and degree of a mapping. In view of this the following natural modifications are proposed.

As in §14 the mappings $f_i(X_i) = Y$ are factored, giving

$f_i(X_i) \;\; = l_i m_i(X_i)$,

$m_i(X_i) = \Sigma_i$, a monotone map from X on to the cactoid Σ_i

$l_i(\Sigma_i) \;\; = Y$, a light mapping from Σ_i onto Y.

Suppose that the general true cyclic element of Σ_i is S_i and consider the mapping $m_i(S_i ; X_i) = S_i$. It is understood that X_1 and X_2 are oriented and so by 11.2 it is possible to select an orientation on S_i such that the degree of this monotone mapping will be 1.

Suppose further that there is a homeomorphism $h(\Sigma_1) = \Sigma_2$. If S_1 and S_2 are corresponding 2-spheres then h is said to have degree 1 (-1) on S_1 if the degree of the partial mapping $h(S_1) = S_2$ is 1 (-1). The homeomorphism h is said to have degree 1 (-1) on Σ_1 if Dgr $h(S_1)$ is 1 (-1) on each true cyclic element of Σ_1.

The mapping $f_1(X_1) = Y$ is said to be *positively Kerékjártó equivalent* to the mapping $f_2(X_2) = Y$ if and only if there is a homeomorphism $h(\Sigma_1) = \Sigma_2$ such that 1) $\rho\{l_1(\sigma_1), l_2(h(\sigma_1))\} = 0$ for $\sigma_1 \epsilon \Sigma_1$ and 2) Dgr $h = 1$ (-1) on Σ_1. (Notation: $f_1 K^+ \sim f_2$).

The mapping $f_1(X_1) = Y$ is said to be *negatively Kerékjártó related* to the mapping $f_2(X_2) = Y$ if and only if there is a homeomorphism $h(\Sigma_1) = \Sigma_2$ such that 1) $\rho\{l_1(\sigma_1), l_2(h(\sigma_1))\} = 0$ for $\sigma_1 \epsilon \Sigma_1$ and 2) Dgr $h = -1$ on Σ_1. (Notation: $f_1K^- \quad f_2$).

Finally the mapping $f_1(X_1) = Y$ is said to be *strongly Kerékjártó equivalent* to the mapping $f_2(X_2) = Y$ if and only if $f_1K^+ \quad f_2$ or $f_1K^- \quad f_2$. (Notation: $f_1K^* \sim f_2$).

There is no trouble in seeing that the first relation is an equivalence, indeed, the following statement is almost as easy to verify.

THEOREM 15.1. *The relation $K^* \sim$ is an equivalence.*

16. The characterization theorems

We have already stated that these modified Kerékjártó equivalences fill the gap left by Kerékjártó and Morrey. The facts are these.

THEOREM 16.1. *If $f_1K^+ \sim f_2$ then $f_1F^+ \sim f_2$, and conversely.*

THEOREM 16.2. *If $f_1K^- \quad f_2$ then $f_1F^- \quad f_2$, and conversely.*

THEOREM 16.3. *If $f_1K^* \sim f_2$ then $f_1F \sim f_2$, and conversely.*

The last theorem is a logical consequence of the first two and so we shall concentrate on these. As a matter of fact we can carry the proofs simultaneously, as they differ only in regard to comments on the degree.

It will be useful to deal with an equivalent form of the Kerékjártó relations. Suppose that the 2-spheres of Σ_i are oriented in the standard way so that Dgr $m_i(S_i\,; X) = 1$ for each S_i of Σ_i, $i = 1, 2$.

If $f_1K^+ \sim f_2$ ($f_1K^- \quad f_2$) we shall be interested in the factorization $f_1(X_1) = l_2hm_1(X)$. Now $hm_1(X) \equiv m_1^*(X)$ is a monotone map from X_1 onto Σ_2 and since Dgr $h = 1$ (-1), Dgr $m_1^*(S_2\,; X_1) = 1$ (-1) for each S_2 of Σ_2. Using the notation l for l_2 and Σ for Σ_2, we have $f_1(X_1) = lm_1^*(X)$ and $f_2(X_2) = lm_2(X_2)$, where Dgr $m_1^*(S; X_1) = 1$ (-1) for each S of Σ. Now let us use the notation m_1 for m_1^*.

$$\left.\begin{array}{lll} \text{Hence} & f_1(X_1) = lm_1(X), \quad m_1(X) = \Sigma, \quad l(\Sigma) = Y, \\ & f_2(X_2) = lm_2(X), \quad m_2(X) = \Sigma, \quad l(\Sigma) = Y, \\ \text{and} & \text{Dgr } m_1(S; X) = 1\ (-1), \text{ while} \\ & \text{Dgr } m\ (S; X) = 1 \text{ for every } S \text{ of } \Sigma. \end{array}\right\} \ldots\ldots\ldots (\text{IV}, 1)$$

Conversely, if two mappings $f_1(X_1)$ and $f_2(X_2)$ satisfy the conditions of (IV, 1) then $f_1K^+ \sim f_2$ ($f_1K^- \quad f_2$).

We are now in a position to state

LEMMA 16.4. *If two mappings $f_1(X_1)$ and $f_2(X_2)$ satisfy the conditions of* (IV, 1) *and $\epsilon > 0$, then there are mappings $f_1^*(X_1)$ and $f_2^*(X_2)$ such that*

$$\left.\begin{array}{l} 1)\ f_i^*(X_i) = l_i^* m_i^*(X_i), \textit{ where } m_i^*(X_i) = \Sigma_i^* \textit{ is monotone, and} \\ \quad l_i^*(\Sigma^*) = Y \textit{ is light};\ i = 1, 2. \\ 2)\ \text{Dgr } m_1^*(S^*; X_1) = 1\ (-1) \textit{ for each } S^* \textit{ of } \Sigma^*. \\ 3)\ \text{Dgr } m_2^*(S^*; X) = 1 \textit{ for each } S^* \textit{ of } \Sigma^*. \\ 4)\ \rho\{f_i(x^i), f_i^*(x^i)\} < \epsilon,\ i = 1, 2. \\ 5)\ \Sigma^* \textit{ is a cluster.} \end{array}\right\} \ldots (\text{IV}, 2)$$

PROOF. In obtaining the new, or refined mappings it will be necessary to use the fact that all the maps are not only continuous but uniformly so, a fact which follows from the compactness of the spaces under consideration. Hence we can assert that if $4\eta = \epsilon > 0$ there is a δ such that $\rho\{\sigma_1, \sigma_2\} < \delta$ implies $\rho\{l(\sigma_1), l(\sigma_2)\} < \eta$.

The theorem is trivial if Σ is degenerate.

We now turn to the fundamental structure theorem (8.6) for Peano spaces and make our first application of uniform continuity. Relative to the space Σ there is a sequence of cyclic chains $C(p_n, q_n)$ such that

1) $\bigcup_1^n C(p_i, q_i) \equiv A_n$ is an A-set for every n.
2) $A_n \cdot C(p_{n+1}, q_{n+1}) = q_{n+1}$ for every n.
3) $\bigcup_1^\infty A_n \equiv H$ is dense in Σ.
4) If d_n is the supremum of the diameters of the components of $\Sigma - A_n$, then $d_n \to 0$.

FIRST MODIFICATION. It is possible to select an n such that $d_n < \delta$. Hence each component of $\Sigma - A_n$ has a diameter $<\delta$, and this guarantees that the image under l of each of these components has a diameter $<\eta$. For convenience we shall employ the notation A for A_n.

The mapping $m_i(A; X_i)$, that is $m_i(X_i)$ followed by the retraction of Σ onto A yields $g_i(X_i) = lm_i(A; X_i)$ where the map l is now restricted to A. We assert that

$$\rho\{g_i(x^i), f_i(x^i)\} < \eta, \qquad x^i \in X_i, \qquad i = 1, 2, \ldots \qquad \text{(IV, 3)}$$

SECOND MODIFICATION. In general there will be an infinite number of 2-spheres in A, on the other hand by 8.4 we know that only a finite number of these can have diameters $\geqq \delta$. Delete from the 2-spheres of A every one which contains a point of the set $(p_1, \cdots, p_n, q_1, \cdots, q_n)$, and those whose diameters are $\geqq \delta$. Then the infinite collection Φ of 2-spheres of A remaining has the following fortunate properties.

1) Each 2-sphere of Φ belongs to $C(p_i, q_i)$ for exactly one index i.
2) Each 2-sphere of Φ contains exactly two of the cut points of A.

Now examine the general 2-sphere S of the collection Φ. Since it has two cut points uniquely determined on it, we can map it homeomorphically onto a geometric sphere so that one of the cut points is taken into the north pole, the other into the south pole. This enables us to speak of the lines of latitude and the meridians on S in the obvious way. A mapping of S onto a meridian of S is now to be defined. A definite meridian is chosen. A general point on S lies on some latitude. This latitude cuts the meridian at precisely one point. It is this point of intersection which is designated as the image of the general point on S. The mapping is clearly continuous, and, in fact, monotone.

If we apply this mapping to each 2-sphere of A in the infinite collection Φ the set A is mapped onto a subset A^* which is also an A-set relative to itself. Designate this mapping by $\phi(A) = A^*$, the mapping $\phi m_i(A; X_i) = A^*$ by

$\mu_i(X_i) = A^*$. Both of these mappings are monotone. Define $k_i(X_i) = l\mu_i(X_i)$ where l is of course restricted to the set A^*. We assert that

$$\rho\{k_i(x^i), g_i(x^i)\} < \eta, \qquad x^i \in X_i, \qquad i = 1, 2 \ldots\ldots\ldots \text{(IV, 4)}$$

We remark that Dgr $\mu_1(S; X_1) = 1\ (-1)$ and Dgr $\mu_2(S; X_2) = 1$ for each S of A^*.

FINAL MODIFICATION. Let $p \in E$ if p is a cut point on a 2-sphere of A^* or if p is a point of the collection $(p_1, \cdots, p_n, q_1, \cdots, q_n)$. The set E is finite. Consider the components of $A - E$. They are finite in number and will be of two types: there will be 2-spheres punctured at a finite number of points; there will be open arcs. For a moment we will be interested in the arcs.

Each open arc has the property that the removal of a certain finite number of its points breaks it into a finite number of components each of diameter $<\delta$. It is possible to perform this operation in each of the open arcs so as to obtain a finite set E^* with the property that: 1) $E^* \supset E$; 2) the components of $A^* - E^*$ consist of precisely those mutilated 2-spheres which are components of $A - E$ and a finite number of open arcs α each of diameter $<\delta$; 3) the end points of α are mapped into distinct points by l.

The mapping $\mu_i(X_i) = A^*$ defined in the second modification is monotone and so each of the inverse sets $\mu_i^{-1}(a)$, $a \in A^*$ is connected, $i = 1, 2$. Consider the upper semi-continuous collection Σ^* consisting of the sets $\mu_1^{-1}(a)$ for a in no α and single points elsewhere. This collection is topologized to yield a cactoid which will be denoted by Σ^* and the related map $m_1^*(X_1) = \Sigma^*$ is defined as usual. This map is of course monotone.

About this construction we make several remarks which can be proved by standard methods.

First, the space Σ^* is a cluster.

Second, the space A^* can be imbedded in Σ^*. Hence it is possible for us to consider $A^* \subset \Sigma^*$.

If we think of $A^* \subset \Sigma^*$ then some of the 2-spheres of Σ^* are already oriented by virtue of the fact that they are also 2-spheres of A^*. For these 2-spheres S^* we know that Dgr $m_1^*(S^*; X) = 1\ (-1)$. Orient the other 2-spheres so that the above requirement as to degree is uniform.

We now wish to define a mapping $m_2^*(X_2) = \Sigma^*$. If S^* is a true cyclic element n A^* then m_2 is defined to be μ_2. Consider a 2-sphere S^* of Σ^* which is not a 2-sphere of A^*. Then S^* has exactly two cut points a and b and a unique arc α of A^* joins them in S^*. There is a mapping of degree 1 from the set $\mu_2^{-1}(\alpha)$ onto S^* which agrees with μ_2 on $\mu_2^{-1}(a)$ and $\mu_2^{-1}(b)$. Do this for each S^* of Σ^* not in A^*. We now obtain a composite mapping $m_2^*(X_2) = \Sigma^*$ such that Dgr $m_2(S^*; X_2) = 1$ for each S^* of Σ^*. It is of course necessary to show that this composite mapping is continuous but this is not difficult.

We pick up the thread of argument in connection with the mapping $l(A^*) \subset Y$ which has meaning only on a subset of Σ^*. Ignore l except on the 2-spheres of $A^* \subset \Sigma^*$, and the points of E^*. On this set define l^* to be l. Consider one of

the arcs α. This lies in some S^* of Σ^* not in A^*. If α has end points a and b then $l^*(a)$ and $l^*(b)$ have been defined. Define $l^*(S^*)$ to be a topological extension onto the 2-sphere in E_3 with $l^*(a)$ and $l^*(b)$ as poles. This construction will work since $l^*(a) \neq l^*(b)$ by our choice of E^*. Proceeding in this fashion we obtain a light mapping $l^*(\Sigma^*) = Y^*$.

(Actually the lightness of this map is important only for esthetic reasons.) Now consider $m_i^*(X_i) = \Sigma^*$ and $l^*(\Sigma^*) = Y^*$, $i = 1, 2$.

Define $f_i^*(X_i) = l^*m_i^*(X_i)$, $i = 1, 2$. We assert that

$$\rho\{f_i^*(x^i), k_i(x^i)\} < 2\eta, \qquad x^i \in X_i, \qquad i = 1, 2. \qquad \text{(IV, 5)}$$

If $x^i \in \mu_i^{-1}(S)$ where S is a true cyclic element of A^*, or if $x \in \mu_i^{-1}(E^*)$ then the distance above is zero. If $x^i \in \mu_i^{-1}(\alpha)$ where the end points of α are a and b then l^* and l agree on a. On the other hand $f_i^*(x^i)$ is within η of $l^*(a)$ and $f_i(x^i)$ within η of $l(a)$.

We can now state that

$$\rho\{f_i^*(x^i), f_i(x^i) < 4\eta = \epsilon \quad \text{for} \quad x^i \in X_i, \qquad i = 1, 2,$$

since this result follows from (IV, 3, 4 and 5).

Moreover, Dgr $m_1^*(S^*; X_1) = 1\ (-1)$, and $m_2(S^*; X_2) = 1$ for each S^* of Σ^*. Hence, in view of the fact that Σ^* is a cluster we have obtained the mappings $f_i^*(X_i)$, $i = 1, 2$ held up as our objective in the lemma.

We make an immediate application of this lemma together with the matching Theorem 12.1 to prove a portion of the Theorems 16.1 and 16.2. There is a ζ such that if $\rho\{\sigma_1^*, \sigma_2^*\} < \zeta$, then $\rho\{l^*(\sigma_1^*), l^*\sigma_2^*\} < \eta$.

By the matching theorem there is a homeomorphism $h_\delta(X_1) = X_2$ of degree $1\ (-1)$ such that $\rho\{m_1^*(x^1), m_2^*(h_\delta(x^1))\} < \zeta$ for every $x^1 \in X_1$. Hence $\rho\{f_1^*(x^1), f_2^*(h_\delta(x^1))\} < \eta$, $x^1 \in X_1$. And now by Lemma 16.4 we have $\rho\{f_1(x^1), f_2(h_\delta(x^1))\} < 4\eta + 4\eta + \eta = 9\eta$ for $x^1 \in x_1$. This shows that

$$\boxed{\begin{array}{l} f_1\, K^+ \sim f_2 \to f_1\, F^+ \sim f_2 \\ f_1\, K^- \quad f_2 \to f_1\, F^- \quad f_2 \\ f_1\, K^* \sim f_2 \to f_1\, F \ \ \sim f_2 \end{array}}$$

The last line is a logical consequence of the first two.

The converse of these propositions is, in some respects, simpler, and a portion of the converse is true in great generality. Since there is no added difficulty in proving the more general theorems we make this our goal.

We note that for each i, if X_i is a general Peano space and $f_i(X_i) = Y$ is a mapping onto the Peano space Y, then, though the hyperspace Σ_i associated with the monotone factor m_i of f_i will not necessarily be a cactoid it is obvious what is meant by the notations $f_1 \quad F \sim f_2$ and $f_1 \quad K \sim f_2$. There is of course no question of orientation and so degree is meaningless.

It will be our purpose to show that $f_1 \quad F \sim f_2$ implies $f_1 \quad K \sim f_2$.

We need to show that there is a homeomorphism between Σ_1 and Σ_2 satisfying

a certain condition relative to the light factors l_1 and l_2 of f_1 and f_2. Since the spaces are compact it is sufficient to show that there is a one to one continuous transformation having this added property. (8.1).

THEOREM 16.5. *If $f_1(X_1) = Y$ and $f_2(X_2) = Y$ are mappings from Peano spaces X_1 and X_2 onto Y, and for every $\epsilon > 0$ there is a continuous transformation $g_\epsilon(X_1) = X_2$ such that $\rho\{f_1(x^1), f_2(g_\epsilon(x^1))\} < \epsilon$, $x^1 \in X_1$, then there is a continuous transformation $\gamma(\Sigma_1) = \Sigma_2$ between the associated hyperspaces such that $\rho\{l_1(\sigma_1), l_2(\gamma(\sigma_1)) = 0$, $\sigma_1 \in \Sigma_1$.*

PROOF. Since X_1 is a Peano space there is a dense set of points $b_1, b_2, b_3, \cdots$ in X_1. Suppose that if $\epsilon = 1/n$ we employ the notation $g_n(X_1)$ for the associated transformation.

The space X_2 is compact and so there is a convergent subsequence of $\{g_n(b_1)\}$. Denote this sequence by $\{g_{1n}(b_1)\}$. For precisely the same reason there is a convergent subsequence of $\{g_{1n}(b_2)\}$. Denote it by $\{g_{2n}(b_2)\}$. Proceeding in this fashion we obtain for each k a convergent subsequence $\{g_{kn}(b_k)\}$ of the sequence $\{g_{k-1n}(b_{k-1})\}$. We notice that $\{g_{kn}(b_j)\}$ is convergent for any $j \leqq k$ since a subsequence of a convergent sequence is convergent. Finally, the sequence $\{g_{nn}(b_k)\}$ is convergent for each k.

The mapping $g_{nn}(X_1) = X_2$ occupies a position in the sequence $\{g_n(X_1)\}$ not before the nth term, hence

$$\rho\{f_1(x^1), f_2(g_{nn}(x^1))\} < 1/n, \qquad x^1 \in X_1 \ldots\ldots\ldots\ldots \text{(IV, 6)}$$

The double subscript is discarded and the notation $g_n(X_1)$ is used. For any k the sequence $\{g_n(b_k)\}$ is convergent.

Select any point σ^1 in Σ_1 and consider the continuum $n_1^{-1}(\sigma^1)$. By the definition of Σ_1 it follows that $f_1(m_1^{-1}(\sigma^1))$ is a single point y of Y. For any number $\eta > 0$ let $U(\eta, y)$ denote the set of points in Y whose distance from y is less than η. This is an open set. The fact that f_1 is continuous guarantees that for any positive integer ν the set $f_1^{-1}(U(1/\nu, y)$ is an open set containing $m_1^{-1}(\sigma^1)$. Let $V_\nu(m_1^{-1}(\sigma^1))$ denote the component of $f_1^{-1}(U(1/\nu, y))$ which contains the continuum $m_1^{-1}(\sigma^1)$. It is convenient to shorten the notation to V_ν if no ambiguity can arise. That V_ν is open follows from the fact that X is Peanian. There is some k such that $b_k \in V_\nu$ and so b_k is in the continuum $\bar{V}_\nu$. Hence $g_n(\bar{V}_\nu) \equiv C_{n\nu}$ contains the convergent sequence of points $g_n(b_k)$. By the Janisewski theorem, $\limsup_n g_n(\bar{V}_\nu)$ is a continuum C_ν.

Select any x^2 in $g_n(\bar{V}_\nu)$. Then there is an x^1 such that $x^1 \in V_\nu$ and $g_n(x^1) = x^2$. It follows that

$$\rho\{f_2(x^2), y\} \leqq \rho\{f_2(g_n(x^1), f_1(x_1^1)\} + \rho\{f_1(x^1), y\} \leqq 1/n + 1/\nu.$$

Therefore

$$f_2(C_{n\nu}) \subset U(1/n + 1/\nu, y) \ldots\ldots\ldots\ldots\ldots\ldots \text{(IV, 7)}$$

If $x^2 \in C_\nu$ then for each integer n there is a point $x^2_n \in C_{n\nu}$ and a subsequence $x^2_{nj} \to x^2$. Choose the point x^1_n in $\bar{V}_\nu$ in such a fashion that $g_n(x^1_n) = x^2_n$. Then

$$\begin{aligned}\rho\{f_2(x^2), y\} &\leqq \rho\{f_2(x^2), f_2(x^2_{nj})\} + \rho\{f_2(x^2_{nj}), y\}\\ &\leqq \rho\{f_2(x^2), f_2(x^2_{nj})\} + 1/n_j + 1/\nu.\end{aligned}$$

The expression on the right converges to $1/\nu$ as j goes to infinity. Hence

$$f_2(C_\nu) \subset U(1/\nu, y) \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(IV, 8)}$$

Finally, consider the sequence $\{V_\nu\}$ and note that $V_\nu \supset V_{\nu+1}$ hence $C_{n\nu} = g_n(\bar{V}_\nu) \supset g_n(V_{\nu+1}) = C_{n\nu+1}$. This implies that $C_\nu = \limsup_n C_{n\nu} \supset \limsup_n C_{n\nu+1} \equiv C_{\nu+1}$. Hence we obtain the continuum $C \equiv \bigcap_1^\infty C_\nu$. In virtue of (IV, 8) above it follows that $f_2(C) \subset U(1/\nu, y)$ for each ν. Hence

$$f_2(C) = y \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(IV, 9)}$$

This means that $C \subset f_2^{-1}(y)$ and as C is a continuum there is an inverse set $m_2^{-1}(\sigma^2)$ containing C. Define

$$\gamma(\sigma^1) = \sigma^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(IV, 10)}$$

As an immediate remark we note that $l_1(\sigma^1) = l_2(\gamma(\sigma^1)) = y$ hence we need only to check the statement that $\gamma(\Sigma_1) = \Sigma_2$ and the continuity of γ to prove the theorem.

First we prove that $\gamma(\Sigma_1) = \Sigma_2$.

Select any σ^2, and let $y = f_2(\sigma^2)$. In exactly the same manner as V_ν was selected relative to $m_1^{-1}(\sigma^1)$ choose the region W_ν containing $m_2^{-1}(\sigma^2)$. We recall that $f_2(W_\nu) \subset U(1/_\nu, y)$.

Now consider $g_n^{-1}(W_\nu) \equiv E_{n\nu}$. This set is not connected, in general, but it is closed since g is continuous.

The space X_1 is compact and so the set $E_\nu = \limsup_n E_{n\nu}$ is not empty. Moreover E_ν is closed.

If $x^1 \in E$ then $\rho\{f_1(x^1), y\} \leqq \rho\{f_1(x^1), f_2(g_n(x^1))\} + \rho\{f_2(g_n(x^1)), y\} \leqq 1/n + 1/\nu$. Hence $f_1(E_{n\nu}) \subset U(1/n + 1/\nu, y)$. Precisely as in (IV, 8) we obtain $f_1(E_\nu) \subset U(1/\nu, y)$. Continuing to follow the pattern we see that $E_{n\nu}, \supset E_{n\nu+1}$ and $E_\nu = \limsup_n E_{n\nu} \supset \limsup E_{n\nu+1} = E_{\nu+1}$. The closed set $E = \bigcap E_\nu$ is such that $f_1(E) \subset U(1/\nu, y)$ for each ν. Hence $f_1(E) = y$.

This means that there is a point σ^1 and a component K of E such that $m_1^{-1}(\sigma^1) \supset K$. We assert that

$$\gamma(\sigma^1) = \sigma^2.$$

This will of course suffice to show that $\gamma(\Sigma_1) = \Sigma_2$.

Suppose that $\gamma(\sigma^1) = \zeta \neq \sigma^2$. Then there are two disjoint regions R_1 and R_2 containing σ^2 and ζ respectively. The inverse sets $m_2^{-1}(R_1)$ and $m_2^{-1}(R_2)$ will also be disjoint regions containing $m_2^{-1}(\sigma^2)$ and $m_2^{-1}(\zeta)$ respectively. Name these regions $U(\sigma^2)$ and $U(\zeta)$.

The notation $V_\nu(m_1^{-1}(\sigma^1))$ is already familiar. We note that $V_\nu \supset K$. As before we obtain the continua $C_{n\nu}$, C_ν and C. By (IV, 10) $C \subset m_2^{-1}(\zeta)$, since $\gamma(\sigma^1) = \zeta$. Hence there is an index μ such that $C_\mu \subset U(\xi)$. Hence there is an integer n_0 with the property that if $n > n_0$

$$C_{n\mu} = g_n(\bar{V}_\nu) \subset U(\zeta) \ldots\ldots\ldots\ldots\ldots\ldots (\text{IV}, 11)$$

On the other hand because of the compactness of X_2 there is an index λ such that $W_\lambda \subset U(\sigma^2)$. Since K is a component of E, the set E intersects $K \subset V_\mu$. Hence there is a sequence of integers $k_1 < k_2 < k_3 < \cdots$ such that $E_{k_j\lambda} = g_{k_j}^{-1}(W_\lambda)$ intersects $\bar{V}_\nu$ for every j. But then $W_\lambda \cdot g_{k_j}(\bar{V}_\mu) \neq 0$ which implies that $U(\sigma^2) \cdot g_{k_j}(\bar{V}_\mu) \neq 0$ for every j. This is contradictory to (IV, 11).

This shows that the mapping γ is from Σ_1 *onto* Σ_2.

We need to check the continuity of γ.

If γ is not continuous then we may suppose the existence of a sequence $\sigma_k^1 \to \sigma^1$ such that $\gamma(\sigma_k^1) \to \zeta \neq \gamma(\sigma^1) = \sigma^2$.

Let $U(\zeta)$ and $U(\sigma^2)$ have the same meaning as above. We recall that the regions are disjoint.

As in the case of (IV, 11) we obtain

$$g_n(\bar{V}_\mu) \subset U(\sigma^2) \ldots\ldots\ldots\ldots\ldots\ldots (\text{IV}, 12)$$

On the other hand $\gamma(\sigma_k^1) = \sigma_k^2 \to \zeta$ and so there is an integer k such that both $m_2^{-1}(\sigma_k^2) \subset U(\zeta)$ and $m_1^{-1}(\sigma_k^1) \subset V_\mu(m_1^{-1}(\sigma^1))$. By (IV, 10) we know that the continuum $m_2^{-1}(\sigma_k^2)$ contains the continuum $\bigcap_\nu \limsup_n g_n(\bar{V}_\nu(m_1^{-1}(\sigma_k^1)))$. Hence for all but a finite number of n we have $g_n(\bar{V}_\nu(m_1^{-1}(\sigma_k^1))) \subset U(\zeta)$. Hence $g_n(m_1^{-1}(\sigma_k^1)) \subset U(\zeta)$ for all but a finite number of n. And now, since $m_1^{-1}(\sigma_k^1) \subset V_\mu(m_1^{-1}(\sigma^1))$ by our choice of k, we see that $g_n(\bar{V}_\mu)$. $U(\zeta) \neq 0$ in flat contradiction to (IV, 12).

This completes the proof of the final part of the theorem.

Using the terminology of 16.5 we have the following

COROLLARY 16.6. *If* $\sigma^2 = \gamma(\sigma^1)$ *then* $m_2^{-1}(\sigma^2) \supset \limsup_n g_n(m_1^{-1}(\sigma^1))$.

PROOF. As in 16.5 we have $V_\nu(m_1^{-1}(\sigma^1))$, from which $C_{n\nu} = g_n(\bar{V}_\nu) \supset g_n(m_1^{-1}(\sigma^1))$. Therefore $C_\nu = \limsup_n g_n(\bar{V}_\nu) \supset \limsup_n g_n(m_1^{-1}(\sigma^1))$ for each ν. (The set on the right is probably not a continuum). But $m_2^{-1}(\sigma^2) \supset C \supset \limsup g_n(m_1^{-1}(\sigma^1))$, as can be seen from (IV, 10).

COROLLARY 16.7. *If each of the mappings* $g_n(X_1) = X_2$ *is a homeomorphism, then* $\gamma(\Sigma_1) = \Sigma_2$ *is also a homeomorphism.*

PROOF. Since Σ_1 is compact it is sufficient to show that γ is one to one, (8.1). Suppose that there are three points σ_1^1, σ_2^1, and σ^2 such that $\gamma(\sigma_1^1) = \sigma^2 = \gamma(\sigma_2^1)$. If $l_2(\sigma^2) = y$ let $W_\nu(m_2^{-1}(\sigma^2))$ be defined as in 16.5. There is an n_0 such that if $n > n_0$ then $g_n(m_1^{-1}(\sigma_1^1))$ and $g_n(m_1^{-1}(\sigma_2^1))$ lie in W_ν. Thus the set $E_{n\nu} = g_n(W_\nu) \supset m_1^{-1}(\sigma_1^1) + m_1^{-1}(\sigma_2^1)$ and is a *continuum* since g_n is a homeomorphism. Hence by the Janisewski theorem $E_\nu = \limsup E_{n\nu} \supset m_1^{-1}(\sigma_1^1) + m_1^{-1}(\sigma_2^1)$ and is a continuum. Finally $E = \bigcap E_\nu \supset m_1^{-1}(\sigma_1^1) + m_1^{-1}(\sigma_2^1)$ and is a continuum.

On the other hand, a duplication of the argument used in showing that the mapping γ is from Σ_1 *onto* Σ_2 shows that $f_1(E) = v$. Hence $E = m_1^{-1}(\sigma_1^1) = m_1^{-1}(\sigma_2^1)$.

We have thus proved

THEOREM 16.8. *If* $f_1 \quad F \sim f_2$, *then* $f_1 \quad K \sim f_2$.

This is the theorem which Kerékjártó stated for 2-spheres.

With the notation of 16.5 and 16.6 we have

THEOREM 16.9. *The sequence* $\{m_2(g_n(x^1))\}$ *converges uniformly to* $\gamma m_1(x^1)$.

PROOF. Suppose G is any region on Σ_1. Let $l_1(\bar{G}) = K$, then by the definition of γ we have $l_2\gamma(\bar{G}) = K$.

Suppose $x^2 \in \limsup g_n m_1^{-1}(\bar{G})$ then there is for each n a point $x_n^2 \in g_n m_1^{-1}(\bar{G})$ and a subsequence $x_{n_j}^2 \to x^2$. Select a point $x_n^1 \in m_1^{-1}(\bar{G})$ such that $g_n(x_n^1) = x_n^2$. We may suppose that $x_{n_j}^1 \to x^1$. Hence $\rho\{f_1(x_{n_j}^1), f_2(x_{n_j}^2)\} < 1/n_j$ and $f_1(x^1) = f_2(x^2)$. But $f_1(x^1) = l_1 m_1(x^1) \in l_1(\bar{G}) = K$. Therefore $f_2(x^2) \in K$. Hence

$$f_2(\limsup_n g_n m_1^{-1}(\bar{G})) \subset K \qquad \text{(IV, 13)}$$

On the other hand G is open and so $m_1^{-1}(\bar{G})$ contains one of the everywhere dense set of points $b_1, b_2, b_3, \cdots$ of 16.5. Hence $\limsup_n g_n m_1^{-1}(\bar{G})$ is a continuum and must lie in some component K^2 of $f_2^{-1}(K)$.

Suppose $\sigma^1 \in \bar{G}$ and $\sigma^2 = \gamma(\sigma^1)$, then by Corollary 16.6 we know that lim sup $g_n m_1^{-1}(\sigma^1) \subset m_2^{-1}(\sigma^2)$. Therefore, lim sup $g_n m_1^{-1}(\bar{G})$. $m_2^{-1}(\gamma(\bar{G})) \neq 0$. Since $m_2^{-1}\gamma(G)$ is a continuum, $\limsup_n g_n m_1^{-1}(\bar{G})$ and $m_2^{-1}(\gamma(\bar{G}))$ lie in the same component K^2 of $f_2^{-1}(K)$.

Now consider the component C^2 of $l_2^{-1}(K)$ which contains $\gamma(\bar{G})$. Such a component exists since $\gamma(\bar{G})$ is a continuum and $l_2(\gamma(\bar{G})) = K$. We notice that $m_2^{-1}(C^2) = K^2$. If the diameter of C^2 is less than some number ϵ then there is also a region G^2 containing C^2 and of diameter $< \epsilon$. Now $\limsup_n g_n m_1^{-1}(G) \subset \limsup_n g_n m_1^{-1}(\bar{G}) \subset K^2 = m_2^{-1}(C^2) \subset m_2^{-1}(G^2)$. Hence there is an integer n_0 such that if $n > n_0$

$$g_n m_1^{-1}(G) \subset m_2^{-1}(G^2) \qquad \text{(IV, 14)}$$

This is the key step in the argument.

The map $l_2(\Sigma_2)$ is light and so by 8.3 there is a δ such that if K is a continuum of diameter $< \delta$ each component of $l_2^{-1}(K)$ is of diameter $< \epsilon$.

The space Σ_1 is the sum of a finite number of regions $G_1, \cdots, G_k$ each of diameter $< \epsilon$ and each having an image under l_1 of diameter $< \delta$. If $l_1(\bar{G}_i) = K_i$ then $d(K_i) < \delta$. Hence each component of $l_2^{-1}(K_i)$ has diameter $< \epsilon$. That component C_i^2 containing $\gamma(\bar{G}_i)$ is contained in a region G_i^2 of diameter $< \epsilon$. Since there are only a finite number of sets in the collection $G_1, \cdots, G_n$ there is an n_0 such that if $n > n_0$

$$g_n m_1^{-1}(G_i) \subset m_2^{-1}(G_i^2), \qquad \text{for } i = 1, \cdots, k.$$

Select any $x^1 \epsilon X_1$. There is a G_i such that $x^1 \epsilon m_1^{-1}(G_i)$. The set G_i^2 contains $\gamma(\bar{G}_i)$ and so $\gamma(m_1(x^1))$. If $n > n_0$ then $g_n(x^1) \epsilon g_n m_1^{-1}(G_i) \subset m_2^{-1}(G_i^2)$. Therefore

$$m_2 g_n(x^1) \epsilon G_i^2 .$$

Hence $\rho\{m_2 g_n(x^1), \gamma m_1(x^1)\} < d(G_i^2) < \epsilon$, and the theorem is proved.

Suppose that S is a 2-sphere of Σ_1 and $\gamma(S^1) = S^2$. For convenience let $m_1(S^1, X_1) \equiv \mu_1(X_1) = S^1$ and $m_2(S^2, X_2) \equiv \mu_2(X_2) = S^2$.

Corollary 16.10. *The sequence $\{\mu_2 g_n(x^1)\}$ converges uniformly to $\gamma\mu_1(x^1)$.*

Proof. For any $\epsilon = 3\eta > 0$ there are at most a finite number of components $K_1^2, \cdots, K_s^2$ of $\Sigma_2 - S^2$ which have diameters $\geqq \eta$. Consider K_j^2 and $a_j^2 = F(K_j^2)$, the frontier of K_j^2. There is a minimum distance 3ζ between points of the set $a_1^2, \cdots, a_s^2$. On S^2 select a connected neighborhood R_j^2 of a_j^2 (open with respect to S^2) with diameter less than ζ and also less than η. Add to R_j^2 all the components of $\Sigma_2 - S^2$ with frontier points in R_j^2. By a standard theorem of the cyclic element theory this new set G_j^2 is a connected open set containing R_j^2. No point of G_j^2 is at a distance greater than 2η from some point of K_j^2. These open sets are disjoint.

Let $\gamma^{-1}(K_j^2) = K_j^1$ and $m_1^{-1}(K_j^1) = K_j$. By theorem 16.9 we know that $m_2 g_n(x^1) \rightrightarrows \gamma m_1(x^1)$ and hence there is an n_0 such that if $n > n_0$ then $m_2 g_n(K_j) \subset G_j^2$. Suppose n_0 is larger than any of the integers $n_1, \cdots, n_s$ and in addition has the property that $\rho\{m_2 g_n(x^1), \gamma m_1(x^1)\} < \eta$ for $n > n_0$ and $x^1 \epsilon X_1$.

Now if $n > n_0$ then we shall show that $\rho\{\mu_2 g_n(x^1), \gamma\mu_1(x^1)\} < 3\eta$.

If $x^1 \epsilon K_j$ then $\gamma\mu_1(x^1) = a_j^2$ and $\mu_2 g_n(x^1) \epsilon R_j^2$. Hence the above distance is less than n. If $x^1 \notin \bigcup K_j$ then $\rho\{\gamma\mu_1(x^1), \gamma m_1(x^1)\} < \eta$ and $\rho\{\mu_2 g_n(x^1), m_2 g_n(x^1)\} > \eta$ hence $\rho\{\mu_2 g_n(x^1), \gamma\mu_1(x^1)\} < 2\eta + \rho\{m_2 g_n(x^1), \gamma m_1(x^1)\} < 3\eta = \epsilon$. This proves the corollary.

Now suppose that X_1 and X_2 are oriented 2-spheres and each true cyclic element S^1 of Σ_1 and S^2 of Σ_2 is so oriented that Dgr $m_1(S^1; X_2) = 1$ and Dgr $m_2(S^2; X_2) = 1$.

In view of 10.6 we have at once

Corollary 16.11. *For large enough n the degree of the partial mapping $\gamma(S^1) = S^2$ for each S^1 of Σ_1 is the same as the degree of $g_n(X_1) = X_2$.*

Theorem 16.12. *If $f_1 \quad F^+ \sim f_2$ then $f_1 \quad K^+ \sim f$; if $f_1 \quad F^- \quad f_2$ then $f_1 \quad K^- \quad f_2$.*

Proof. The result follows by virtue of 11.2, 16.5 and 16.11.

It is now possible to tabulate the results of this chapter in the following diagram

$$\boxed{\begin{array}{lcl}
f_1F^+ \sim f_2 & \leftrightarrow & f_1K^+ \sim f_2 \\
f_1F^- \quad\; f_2 & \leftrightarrow & f_1K^- \quad\; f_2 \\
f_1F \;\; \sim f_2 & \leftrightarrow & f_1K^* \sim f_2 \\
f_1F \;\; \sim f_2 & \rightarrow & f_1K \;\; \sim f_2 \\
f_1F \;\; \sim f_2 & \nleftarrow & f_1K \;\; \sim f_2
\end{array}}$$

,

CHAPTER V

TOPOLOGICAL APPLICATIONS

The results of Chapter IV are now applied to the fundamental problem mentioned in §7, and to other questions concerning representation.

17. Representations

It has already been mentioned that if the equivalence relation used is Fréchet equivalence, then the equivalence classes obtained will be called Fréchet surfaces. In the event we use positive Fréchet equivalence the resulting classes will be called oriented Fréchet surfaces. The notation for the first class will be $[f]$ whereas for the second we will use the symbol $[f]^+$.

Suppose that $[f]^+$ is an oriented Fréchet surface with a single known representation $f(X) = Y$. The factorization for f is $lm(X)$, $m(X) = \Sigma$, $l(\Sigma) = Y$. Suppose that X is oriented, then this induces an orientation on each 2-sphere of Σ; that is, the orientation such that Dgr $m(S; X) = 1$ for each 2-sphere S of Σ. Suppose that we have a mapping $m_1(X_1) = \Sigma$ from an oriented 2-sphere X_1 onto Σ having the following properties: 1) m_1 is monotone, and 2) Dgr $m_1(S; X_1) = 1$ (-1) for each 2-sphere S of Σ. If m has these properties we say that it is monotone and positively (negatively) consistent with m. A mapping m is said to be monotone and consistent with m if it is monotone and either positively or negatively consistent with m.

Consider the mapping $f_1(X_1) = lm_1(X_1)$. By Theorem 16.1 (16.2) we know that if m_1 is monotone and positively (negatively) consistent with m then $f_1(X_1)F^+ \sim f(X)$ $(f_1(X_1)F^- \sim f(X))$. Moreover, if $f_1(X_1)F^+ \sim f(X)$ $(f_1(X_1)F^- \sim f(X))$ then there is a monotone mapping $m_1(X_1) = \Sigma$ which is positively (negatively) consistent with m such that $f_1(X_1) = lm_1(X)$. In view of these remarks we have the

FUNDAMENTAL STRUCTURE THEOREM 17.1. If $f(X) = lm(X)$, $m(X) = \Sigma$, is a representation of an oriented Fréchet surface $[f]^+$ (Fréchet surface $[f]$), then the totality of representations of $[f]^+([f])$ is the totality of mappings of the form $f_1(X_1) = lm_1(X_1)$ where $m_1(X_1) = \Sigma$ is any monotone map positively consistent (consistent) with m.

We do not plan to make any direct topological applications of this theorem in this paper, but it is clear that if we are to obtain representations which enjoy certain desirable properties then the mode of attack is to alter the monotone factor of the given representation accordingly.

A rather useful theorem in many applications is the weaker

THEOREM 17.2. *If $f(X) = lm(X)$, $m(X) = \Sigma$, is a representation of an oriented Fréchet surface $[f]^+$ (Fréchet surface $[f]$), and $\mu(X) = X$ is a monotone mapping of degree 1 (is any monotone mapping) then $f(\mu(X))$ is also a representation of the oriented Fréchet surface $[f]^+$ (Fréchet surface $[f]$).*

CHAPTER VI

AREA

The applications in this chapter are concerned with the problem of area both for oriented and ordinary Fréchet surfaces. We state the case for oriented surfaces; there will be no difficulty in making the usual translation to the case of Fréchet surfaces.

The repeated insertion of the adjective *oriented* becomes laborious and so we drop it on the explicit understanding that all statements are made for oriented surfaces. In conformity with this understanding we write $[f]$ instead of $[f]^+$ as heretofore.

It should also be noted that the general outline of the application is due to Morrey. In spite of the fact that Morrey based his whole argument on a false structure theorem, his final theorems are, in the main, correctly stated.

18. Preliminary concepts

Without attempting an exposition of the subject—here we can consult Radó [7]—the situation for oriented Fréchet surfaces is roughly this. The collection of all such surfaces is topologized and in the resulting space there is selected a favored class of surfaces called oriented polyhedra. Topologically this favored class is important because it is dense; analytically it is important because each polyhedron $[p]$ can be assigned a non-negative number $E[p]$—known as the elementary area in a perfectly natural manner. Moreover, $E[p]$ is a lower semi-continuous function on the class of polyhedra. Explicitly, if $[p_n] \to [p]$ in the topology and all surfaces are polyhedra, then $\liminf E[p_n] \geqq E[p]$.

The Lebesgue area of a surface $[f]$ is now defined by extending the function $E[p]$. Since the polyhedra are dense in the space there is at least one sequence $[p_n] \to [f]$. With this particular sequence there is associated a non-negative number defined to be $\liminf E[p_n]$. (For convenience infinity is spoken of as a number.) The infimum of the numbers $\liminf E[p_n]$ is defined to be the Lebesgue area of $[f]$.

$$L[f] = \inf_{[p_n]\to[p]} \liminf_{n} E[p_n] \text{.....................(VI, I)}$$

The topologization of the collection of surfaces is afforded by the oriented Fréchet metric defined by McShane [4]. An equivalent topologization is the following $[f_n] \to [f]$ if and only if there is a representation $f(X)$ of $[f]$ and for each n a representation $f_n(X)$ of $[f]$ such that $f_n(x) \rightrightarrows f(x),\ x \in X$. (The notation stands for uniform convergence.)

The literature contains several methods of selecting the class of polyhedra, the definition depending on the manner in which the class is to be employed. The different methods all yield the same number for the Lebesgue area.

For the purpose of this discussion a polyhedron is a surface $[p]$ which has a representation of a particular kind now to be defined.

To begin with the range is understood to be a geometrical sphere in E_3. The concept of a triangulation and η-triangulation has already been introduced, (11.1). A transformation $p(X) = Y$ is said to be a *polyhedral mapping* if there is a triangulation T of the oriented 2-sphere X such that for each Δ of T the mapping $p(x)$, $x \in \Delta$ is: 1) a homeomorphism onto a rectilinear triangle $p(\Delta)$ in E_3; or 2) a monotone map onto a segment of a straight line in E_3; or 3) a (monotone) map onto a single point in E_3. The triangulation T is said to be associated with the polyhedral mapping $p(X) = Y$.

A surface $[p]$ is an (oriented) *polyhedron* if it has a representation $p(X) = Y$ which is polyhedral.

Consider a polyhedral mapping $p(X) = Y$ and associated triangulation T. If $\Delta \in T$ then $p(\Delta)$ has an area $E(p(\Delta))$ in the ordinary sense of the word, an area which may, of course, be zero. Let $E(p) = \Sigma E(p(\Delta))$ where the summation is taken over all Δ's in T. This is called the elementary area of the *mapping* $p(X) = Y$. The important thing is that if a polyhedron p has two polyhedral representations $p_1(X_1)$ and $p_2(X_2)$, then $E(p_1) = E(p_2)$. Hence this common value may be designated by $E[p]$ and spoken of as the elementary area of the polyhedron $[p]$.

A triangulation T^* is said to be a *refinement* of the triangulation T if for each Δ^* of T^* there is a Δ of T such that $\Delta^* \subset \Delta$. We notice that if $p(X)$ is a polyhedral mapping with associated triangulation T, and T^* is a refinement of T, then it is not true in general that $p(X)$ is polyhedral with respect to the triangulation T^*.

We need a further remark on the definition of area. It is clear from the definition that for any surface $[f]$ there exists a sequence of polyhedra $\{[p_n]\}$ such that $[p_n] \to [f]$ and also $E[p_n] \to L[f]$. The first limit means that there is a representation $f(X)$ of $[f]$ and a representation $p_n(X)$ of $[p_n]$ such that $p_n(x) \rightrightarrows f(x)$, $x \in X$. Unfortunately there is no a priori guarantee that $p_n(x)$ will be a polyhedral mapping. It is known, however, that given any representation $f(X)$ of a surface $[f]$ from a geometrical sphere X in E_3, there is a sequence of polyhedral mappings $P_n(X)$ and associated triangulations T_n such that

$$\left.\begin{array}{l} 1)\ p_n(x) \rightrightarrows f(x), \quad x \in X. \\ 2)\ E[p_n] \to L[f]. \\ 3)\ |T_n| \to 0 \\ 4)\ \delta(T_n; p_n) = \max\limits_{\Delta \in T_n} d(p_n(\Delta)) \to 0 \end{array}\right\} \ldots\ldots\ldots\ldots\ldots (VI, 2)$$

(For a discussion of the comparable case on a 2-cell, see Youngs [12]).

19. The stretching process

Suppose $p(X) = Y$ is a polyhedral mapping with an associated triangulation T consisting of N triangles. Associated with the triangle Δ on the sphere X is the rectilinear triangle Δ^* whose vertices are the same as the vertices of the

triangle Δ. Consider a rectilinear coordinate system on Δ^* and (ξ, η, ζ) system in E_3, with origin y_0. There is a unique map $\xi = \xi(u, v)$, $\eta = \eta(u, v)$, $\zeta = \zeta(u,v)$, linear in both u and v—expressible in vector notation as $\mathfrak{X} = \mathfrak{X}(u, v)$—defined by the requirement that $\mathfrak{X}(u, v)$ agree with $p(x)$ on the vertices of Δ^*. These vertices are, of course, the only points of Δ^* at which $p(x)$ is defined.

For $R > r > 0$ consider the vector mapping

$$\mathfrak{Y}(u, v) = \frac{R[|\mathfrak{X}(u, v)| - r]}{|\mathfrak{X}(u, v)|[R - r]} \cdot \mathfrak{X}(u, v) \ldots\ldots\ldots\ldots\ldots (\text{VI}, 3)$$

due to Morrey [6], Radó and Reichelderfer [9].

For any integer n consider a triangulation T_n^* of Δ^* with the important property that the vertices of T_n^* divide each side of Δ^* into exactly n equal parts. Suppose $\delta_1^*, \cdots, \delta_k^*$ are the triangles of T_n^*.

There is a unique linear map $\mathfrak{X}_i^*(u, v)$ from δ_i^* compatible with the requirement that $\mathfrak{X}_i^*(u, v)$ agree with $\mathfrak{Y}(u, v)$ at the vertices of δ_i^*. The mapping $\mathfrak{X}^*(u, v)$ defined to be $\mathfrak{X}_i^*(u, v)$ if $(u, v) \epsilon\, \delta_i^*$ is known as quasi-linear, [McShane 4]. The ordinary area of the image triangle $\mathfrak{X}_i^*(\delta_i^*)$ is denoted by $E(\mathfrak{X}_i^*(\delta_i^*))$, and the sum of these numbers, for $i = 1, \cdots, k$, is denoted by $E(\mathfrak{X}^*(\Delta^*))$. The mapping $\mathfrak{Y}(u, v)$ is smooth enough to argue by classical results that for any $\epsilon > 0$ there is an integer $n(\epsilon, \Delta^*)$ such that if $n > n(\epsilon, \Delta^*)$ then there is a triangulation T_n^* of the above type with the property that, for the associated quasi-linear mapping $\mathfrak{X}^*(u, v)$ we have

$$E(\mathfrak{X}^*(\Delta^*)) < I(\mathfrak{X}^*) + \epsilon,$$

where the symbol $I(\mathfrak{X}^*)$ represents the number obtained by an application of the standard integral formula for the area.

A highly interesting result of Morrey [6], Radó and Reichelderfer [9] enables us to assert that

$$I(\mathfrak{Y}^*) \leqq \frac{R}{R - r} \cdot E(p(\Delta))$$

and so

$$E(\mathfrak{Y}^*(\Delta)) < \frac{R}{R - r} \cdot E(p(\Delta)) + \epsilon.$$

For each Δ^* of T there is an associated $n(\epsilon, \Delta^*)$. Select an integer n larger than the maximum of these numbers and consider a triangulation T_n^* for each Δ^* of T. The projection of these triangulations onto X from the center of X yields a refinement T_n of the triangulation T. (It is understood that the original T is fine enough to allow this statement.)

Define a modified mapping $p^*(x)$ on this wise. The point x is the projection of a point (u, v) on some δ_i^* of a triangle Δ^*. Let $p^*(x) = \mathfrak{X}_i^*(u, v)$. The special

choice of triangulations T_n^* relative to the sides of Δ^* guarantees that $p^*(x)$ is continuous over X, in fact, the mapping is polyhedral. Moreover,

$$\left.\begin{array}{l} 1)\ E(p^*(X)) < \dfrac{R}{R-r} E(p(X)) + N\epsilon, \\ 2)\ \rho\{p^*(x), p(x)\} < \delta(T; p) + R,\ x \in X \\ 3)\ \text{If a triangle } \Delta \text{ of } T \text{ is such that} \end{array}\right\} \cdots\cdots\cdots (\text{VI}, 4)$$

$p(\Delta)$ has every point at a distance less than r from y_0, then $p^*(\Delta) = y_0$.

We shall speak of having obtained $p^*(x)$ from $p(x)$ relative to y_0, R, r and ϵ.

20. The additivity of the area

The applications of the stretching process are concerned with a situation in which we have a surface $[f]$ with representation $f(X) = lm(X) = Y$, $m(X) = \Sigma$ and Σ has a cut point c. Then $\Sigma = \Gamma_1 + \Gamma_2$ where both Γ_1 and Γ_2 are open and $\Gamma_1 \cdot \Gamma_2 = 0$. In fact, placing $\bar{\Gamma}_1 = \Sigma_1$ and $\bar{\Gamma}_2 = \Sigma_2$, both Σ_1 and Σ_2 are cactoids and have the single point c in common.

For each i, let l_i be the partial mapping $l(\Sigma_i) \subset Y$ so that we can write $l_i(\Sigma_i) = Y_i \subset Y$. Let $f_i(X) = l_i m(\Sigma_i\,; X)$.

For each true cyclic element S of Σ choose the orientation so that $\text{Dgr}\ m(S; X) = 1$. Let X_1 and X_2 be two oriented geometric spheres in E_3 the point of tangency being d. By a standard argument there is a monotone mapping $m_i(X_i) = \Sigma_i$ such that $m_i(S; X_i)$ has degree 1 for each 2-sphere S in Σ_i, $i = 1, 2$. (This is, but for degree, Theorem 8.8.)

Now consider the sphere X. There is a monotone map $\mu(X) = X_1 + X_2$ which is one to one except that $\mu^{-1}(d)$ is a great circle which we shall call the equator Q of X. Moreover $\mu(X_i\,; X) = X_i$ has degree 1 for each i. Suppose that $\mu^{-1}(X_i) = A_i$. Then $A_1 \cdot A_2 = Q$.

$$\text{The monotone map } m^*(x) = \begin{cases} m_1\mu(x), & x \in A_1 \\ m_2\mu(x), & x \in A_2 \end{cases}$$

is of some importance. Not only is $m^*(X) = \Sigma$ but $m^*(S; X)$ has degree 1 for each 2-sphere S of Σ. By the fundamental structure Theorem 17.1 the mapping $f^*(X) = lm^*(X)$ is also a representation of $[f]$, and so $L[f] = L[f^*]$.

Let $m_i^*(X) = m_i\mu(X_i\,; X)$, and $f_i^*(X) = l_i m_i^*(X)$, $i = 1, 2$. We now have $[f_i] = [f_i^*]$.

THEOREM 20.1. $L[f] = L[f_1] + L[f_2]$.

PROOF. PART I. $L[f^*] \geqq L[f_1^*] + L[f_2^*]$.

From (VI, 2) we know that there is a sequence of polyhedral mappings $p_n(X)$ such that $\rho\{p_n(x), f^*(x)\} < 1/n^2$, $x \in X$, $E[p_n] \rightarrow L[f^*]$ and both $|\,T_n\,|$ and $\delta(T_n\,; p_n)$ are less than $1/n^2$. If $x \in Q$ then $p_n(x) \rightrightarrows f^*(x) = lm^*(x) = l(c) = y_0$. By $U(r)$ we shall understand the set of points in E_3 at a distance less than r from y_0. For any integer k consider $U(1/k^2)$. There is a sequence of integers $n_1 < n_2 < n_3 < \cdots$ such that $p_{n_k}(x) \in U(1/k^2)$ for $x \in Q$. Discard the poly-

hedral mappings not in the sequence $\{p_{n_k}\}$ and rename the sequence $\{p_k\}$. It is clear that $n_k \geqq k$ and so $1/n_k \leqq 1/k$.

Suppose there are N_k triangles in the triangulation T_{n_k} which is now renamed T_k. Let $\epsilon = 1/(k \cdot N_k)$ and using $R = 2/k$, $r = 2/k^2$ obtain $p_k^*(x)$ from $p_k(x)$ relative to y_0, R, r, and ϵ. Then by (VI, 4)

$$E[p_k^*(x)] \leqq E[p_k(x)] \cdot \frac{1}{1 - 1/k} + 1/k$$

$$\begin{aligned} \rho\{p_k^*(x), p_k(x)\} &< \delta(T_k ; p_k) + R \\ &< 1/n_k^2 + 2/k \\ &< 1/k^2 + 2/k. \end{aligned}$$

Moreover, any Δ of T_k having a point in common with Q is such that $p_k(\Delta)$ has a point in $U(1/n_k^2)$ and since its diameter is less than $1/n_k^2 \leqq 1/k^2$ it follows that $p_k(\Delta) \subset U(2/k^2) = U(r)$. This implies that $p_k^*(\Delta) = y_0$.

Let

$$p_{ik}^*(x) = \begin{cases} p_k^*(x) & \text{if } x \in A_i \\ y_0 & \text{if } x \notin A_i \end{cases} \qquad i = 1, 2.$$

For each i, the mapping $p_{ik}^*(x)$ is polyhedral (the associated triangulation is T_k^*). Furthermore $E[p_{1k}^*] + E[p_{2k}^*] = E[p_k^*]$ and $p_{ik}^*(x) \rightrightarrows f_i^*(x)$.

Hence $L[f_i^*] \leqq \liminf E[p_{ik}^*]$, $i = 1, 2$.

Now $L[f^*] = \lim E[p_k] = \lim E[p_k^*] \geqq \liminf E[p_{1k}^*] + \liminf E[p_{2k}^*] \geqq L[f_1^*] + L[f_2^*]$.

Part II: $L[f^*] \leqq L[f_1^*] + L[f_2^*]$.

For each i select a sequence of polyhedral mappings $p_{in}(x)$ with associated triangulations T_{in} such that $\rho\{p_{in}(x), f_i^*(x)\} < 1/n^2$ uniformly in x, $E[p_{in}] \to L[f_i^*]$ and both $| T_{in} |$ and $\delta(T_{in} ; p_{in})$ are $< 1/n^2$. (IV, 2).

For each i repeat the process of part I to obtain $p_{ik}^*(x)$ such that

$$E[p_{ik}] \leqq E[p_{ik}] \cdot \frac{1}{1 - 1/k} + 1/k$$

$$\rho\{p_{ik}^*(x), p_{ik}(x)\} < 1/k^2 + 2/k$$

$$p_{1k}^*(A_2) = y_0, \qquad p_{2k}^*(A_1) = y_0$$

Define

$$p_k^*(x) = p_{ik}^*(x) \quad \text{if } x \in A_i, \qquad i = 1, 2.$$

Now $p_k^*(x)$ is a polyhedral mapping and the associated triangulation is essentially T_{1k} in A_1 and T_{2k} in A_2. A slight modification must be made in those triangles Δ of T_{ik} which have two sides cut by Q.

We have $E[p_k^*] = E[p_{1k}^*] + E[p_{2k}^*]$ and $p_k^*(x) \rightrightarrows f^*(x)$. Hence $L[f^*] \leqq \liminf E[p_k^*] = \lim E[p_{1k}^*] + \lim E[p_{2k}^*] = L[f_1^*] + L[f_2^*]$.

This completes the proof of the theorem.

This theorem is the key to a much more general situation in which we have additivity. Suppose, for example, that the cactoid Σ associated with a surface $[f]$ is a cyclic chain $C(a, b)$. Let α be an arc from a to b and suppose $a_1, \cdots, a_n$ are cut points of $C(a, b)$ in order from a to b along α. Then $C(a_i, a_{i+1}) \equiv C_i$ is a cyclic chain from a_i to a_{i+1} for $i = 0, \cdots, n$; $a = a_0$, $b = a_{n+1}$.

Consider $f_i(X) = lm(C_i; X)$ for $i = 0, \cdots, n$. These mappings represent surfaces $[f_i]$ and we have the

COROLLARY 20.2. $L[f] = L[f_0] + \cdots + L[f_n]$.

PROOF. The argument is by induction on the number of cut points. The theorem is obviously true if there are no cut points. Suppose it is true for any collection of $(n - 1)$ cut points. Consider the collection of cut points $a_1, \cdots, a_{n-1}$ then we have $L[f] = L[f_0] + \cdots + L[f_{n-2}] + L[lm(C(a_{n-1}, b); X)]$. Now to the mapping $lm(C(a_{n-1}, b); X)$ apply the theorem with a_n replacing c. The chain $C(a_{n-1}, b)$ is broken up into two sub-cactoids $C(a_{n-1}, a_n) \equiv C_{n-1}$ and $C(a_n, b) \equiv C_n$. If $f_{n-1} = lm(C_{n-1}; X)$, $f_n = lm(C_n; X)$ then $L[lm(C(a_{n-1}, b); X)] = L[f_{n-1}] + L[f_n]$, and the assertion is proved.

LEMMA 20.3. *If $[f]$ is a surface and the associated cactoid Σ is an arc then* $L[f] = 0$.

PROOF. By 17.1 we can obtain a representation $f(X) = lm(X)$ such that $m^{-1}(\sigma)$ is a line of longitude on X unless σ is an end point of Σ and in this exceptional event $m^{-1}(\sigma)$ is either the north or south pole of X. It is now clear that we can find polyhedral mappings $p_n(X)$ such that $p_n(x) \rightrightarrows f(x)$ and $E[p_n] = 0$.

We return to the case of a surface $[f]$ whose associated cactoid Σ is a single cyclic chain $C(a, b)$. Consider an arc α as before from a to b and let $C^*(a, b)$ denote the modified chain consisting of the arc α together with all the 2-spheres in $C(a, b)$ of diameter $> 1/n$. These are finite in number. Let the cut points of $C(a, b)$ on these 2-spheres be $a_1, \cdots, a_m$. Consider a mapping $\mu_n(\Sigma) = C_n$ which is the identity except on 2-spheres of diameter $\leqq 1/n$, and on each of these is a monotone map of the 2-sphere onto the segment of α it contains leaving the cut points fixed.

Let ${}_nf(X) = l\mu_n m(X)$; the associated cactoid is $C_n(a, b)$. If we denote a by a_0 and b by a_{m+1} then with the subchains $C_n(a_i, a_{i+1})$ we have partial mappings ${}_nf_i$ as in corollary 20.2, for $i = 0, \cdots, m$. Hence

$$L[{}_nf] = L[{}_nf_1] + \cdots + L[{}_nf_m].$$

The mappings ${}_nf_i(X)$ fall into precisely two categories: there are some for which the associated cactoid $C_n(a_i, a_{i+1})$ is a 2-sphere; there are others for which it is an arc. In the second case $L[{}_nf_i] = 0$.

With the original mapping $f(X)$ and cactoid $C(a, b)$ we have the subcactoids $C(a_i, a_{i+1})$ and partial mappings $f_i(X)$ as in corollary 20.2, for $i = 0, \cdots, m$. Hence

$$L[f] = L[f_1] + \cdots + L[f_m].$$

The mappings $f_i(X)$ also fall into precisely two categories: there are some for which the associated cactoid $C(a_i, a_{i+1}) = C_n(a_i, a_{i+1})$ is a 2-sphere; then there are the remaining mappings which we do not characterize in their own right.

For a mapping $f_i(X)$ of the first type $f_i(X) = {}_nf_i(X)$ and so $L[f_i] = L[{}_nf_i]$; for a mapping of the second $L[f_i] \geqq 0 = L[{}_nf_i]$. Hence we can draw the conclusion that $L[f] \geqq L[{}_nf]$. On the other hand it is clear that ${}_nf(x) \rightrightarrows f(x)$ and hence lim inf $L[{}_nf] \geqq L[f]$. These two facts imply that $L[{}_nf] \to L[f]$.

The above discussion shows that if $S_1, \cdots, S_{k_n}$ are the 2-spheres of diameter $>1/n$ in $C(a, b)$ and $g_i(X) = lm(S_i; X)$ then $L[{}_nf] = L[g_1] + \cdots + L[g_{k_n}]$. Hence $L[f] = \lim_{n\to\infty} \{L[g_1] + \cdots + L[g_{k_n}]\}$.

For terminological convenience we shall record the 2-spheres of Σ as S_1, $S_2, S_3, \cdots$ and the mappings $lm(S_i; X)$ by $f_i(X)$, then we have proved the

LEMMA 20.4 $L[f] = \sum_1^\infty L[f_i]$.

We are at last in a position to examine a general mapping $f(X) = lm(X) = Y$, $m(X) = \Sigma$. Let A be any A-set in Σ. The components of $\Sigma - A$ can be arranged in order by size as only a finite number are larger than $1/n$, $n = 1, 2, 3, \cdots$ Suppose such an arrangement is $B_1, B_2, B_3, \cdots$. For each i the set $F(B_i)$ is a single point b_i. Define $f(A; X) = lm(A; X)$, and $\mu_k(\Sigma) \epsilon \Sigma$ to be the mapping which is the identity except on $B_1, \cdots, B_k$. If $\sigma \epsilon B_j$, $1 \leqq j \leqq k$, then $\mu(\sigma) = b_j$. This is a monotone retraction. Let $f_k(X) = l\mu_k m(X)$. It follows that $f_k(x) \rightrightarrows f(A; x)$, and so lim inf $L[f_k(X)] \geqq L[f(A; X)]$.

As a consequence of 20.1 it is true that $L[f] \geqq L[f_1]$ and in general $L[f_k] \geqq L[f_{k+1}]$. Hence $L[f] \geqq L[f(A; X)]$.

Consider a chain decomposition $\{C_i\} \equiv \{C(p_i, q_i)\}$ of Σ guaranteed by 8.6. Using the notation of 8.6 let $lm(A_n; X)$ be denoted by $g_n(X)$. The compactness of the spaces involved assures us that $g_n(x) \rightrightarrows f(x)$. Hence lim inf $L[g_n] \geqq L[f]$. But $L[g_n] \leqq L[f]$. Hence lim $L[g_n] = L[f]$.

As for the mapping $f(C_i; X) = lm(C_i; X)$, an argument using 20.1 shows that $L[g_n] = L[f(C_1; X)] + \cdots + L[f(C_n; X)]$. Hence $L[f] = \sum_1^\infty L[f(C_i; X)]$. (Note the proof and discussion of Corollary 20.2)

If the 2-spheres of Σ are recorded as $S_1, S_2, S_3, \cdots$ and $f_i(X) = lm(S_i; X)$, then using Lemma 20.4 we have a proof for the

ADDITIVITY THEOREM 20.5. $L[f] = \sum_1^\infty L[f_n]$.

In this connection it is interesting to note an elegant result of Radó [8].

THEOREM 20.6. If the cactoid Σ associated with a mapping $f(X)$ is a single 2-sphere, then $L[f] > 0$.

In the event there are 2-spheres in each of the cyclic chains $C(p_i, q_i)$ of Theorem 20.5 then the mappings $g_n(X)$ defined above, which converge uniformly to $f(X)$, are such that

$$L[g_1] < L[g_2] < L[g_3] < \cdots \to L[f].$$

It follows that it is possible to find a sequence of polyhedral mappings $p_n(x) \rightrightarrows f(x)$ such that

$$E[p_1] < E[p_2] < E[p_3] < \cdots \to L[f].$$

This result is mentioned here because of its independent interest. It would be convenient to have the weaker result with $\leqq$ replacing $<$ regardless of the structure of Σ.

A second consequence of the theorem of Radó (20.6) should be mentioned. In conjunction with the Additivity Theorem 20.5 we have

THEOREM 20.7. *$L[f] = 0$ if and only if the associated cactoid Σ is a dendrite; i.e., has no 2-spheres.*

For a direct proof of this result see Radó [8]. The theorem carries an air of fitness in view of the slender structure of a dendrite.

It should perhaps be mentioned again, as in the beginning of this chapter, that these theorems are equally true for ordinary Fréchet surfaces. In addition, it is clear that, if $f(X)$ is any mapping, then $L[f^{+}] = L[f]$.

A final application of the additivity theorem shows that the following statement is true.

THEOREM 20.8. *If $f_1K \sim f_2$ then $L[f_1] = L[f_2]$.*

APPENDIX

So far the theory developed has been exclusively for surfaces of the type of a 2-sphere. From the analytic point of view it is surfaces of the type of a 2-cell which are more convenient. In this appendix some comments are made concerning such surfaces.

Without bothering to redefine all the concepts explicitly it is clear that such terms as positively Fréchet equivalent and positively Kerékjártó equivalent, to mention two, carry over at once. In fact, the theorems are equally true for surfaces of the type of a 2-cell.

There are several ways of checking this last statement. Any surface of the type of a 2-cell gives rise to a surface of the type of a 2-sphere by a "doubling up" process. Explicitly, we may suppose the surface $[f]$ of the type of a 2-cell has a representation f whose range is the northern hemisphere of a geometrical sphere. If ϕ is the mapping reflecting the southern hemisphere about the equatorial plane onto the northern hemisphere, then $f\phi$ is a mapping which together with f gives a composite mapping which is a representation of some surface $[f^*]$ of the type of a 2-sphere.

It is not difficult to see that $L[f^*] = 2L[f]$. Moreover, the hyperspace Σ is divided symmetrically into two hemicactoids Σ_1 and Σ_2 in an obvious way. In this fashion it is an easy exercise to arrive at an additivity theorem which is the analog of 20.5.

In regard to the doubling up technique one should be warned that if f_1 and f_2 are two maps from 2-cells, and f_1^* and f_2^* are the results of doubling up, then though $f_1F \sim f_2$ (for example) implies $f_1^*F \sim f_2^*$, the converse is *false*. This imposes some limitations on this mode of attack when applied to the topological aspects of the situation. Indeed, for the topological analogs a satisfactory technique appears to be a use of the characterization of hemicactoids as in Whyburn [10], together with a somewhat stronger approximation and matching

theorem. One has to exercise caution at the edge of the 2-cell. Though this plan of attack carries through it would be pleasing to reduce the 2-cell case to the argument for the 2-sphere by employing a simple artifice.

It will have been noticed that the fundamental structure Theorem 17.1 is essentially used in the application to area only in obtaining a special representation for the surface. (We obtained f^* from f and f_i^* from f_i, $i = 1, 2$.) It seems likely that a direct approach can be found.

PURDUE UNIVERSITY.

BIBLIOGRAPHY

[1] ADKISSON, V. W. AND MACLANE, SAUNDERS. *Extending maps of plane Peano continua.* Duke Math. Journ. 6 (1940), pp. 216–228.

[2] ALEXANDROFF, P. AND HOPF, H. *Topologie*, vol. 1, Berlin, 1935.

[3] KERÉKJÁRTÓ, B. v. *Involutions et surfaces continues*, Szeged, Acta Litterarum ac Scientiarum, 3 (1927), pp. 49–67.

[4] MCSHANE, E. J. *On the semi-continuity of double integrals in the calculus of variations.* Ann. of Math. 33 (1932), pp. 460–484.

[5] MORREY, C. B. *The topology of (path) surfaces.* Amer. Journ. of Math. 57 (1935), pp. 17–50.

[6] ——— *An analytic characterization of surfaces of finite Lebesgue area, part* II. Amer. Journ. of Math. 58 (1936), pp. 313–322.

[7] RADÓ, T. *On the semi-continuity of double integrals in parametric form.* Trans. Amer. Math. Soc. 51 (1942), pp. 336–361.

[8] ——— *On continuous path surfaces of zero area.* Ann. of Math. 44 (1943), pp. 173–191

[9] RADÓ, T. AND REICHELDERFER, P. *On a stretching process for surfaces.* Amer. Journ. of Math. 61 (1939), pp. 645–650.

[10] WHYBURN, G. T. *Analytic Topology.* Amer. Math. Soc. Colloquium Series, vol. 28, New York, 1942.

[11] WHYBURN, G. T. AND KURATOWSKI, C. *Sur les éléments cycliques et lears applications,* Fund. Math. 16 (1930), pp. 305–331.

[12] YOUNGS, J. W. T. *The additivity of the Lebesgue area.* Bull. Amer. Math. Soc. 49 (1943), pp. 779–784.

ANNALS OF MATHEMATICS
Vol. 45, No. 4, October, 1944

IN GENERAL A MEASURE PRESERVING TRANSFORMATION IS MIXING

BY PAUL R. HALMOS

(Received November 13, 1943)

In this note I continue the study of topological properties of the group of measure preserving transformations begun in an earlier paper.[1] Using the methods and results of that paper I present the first proof of the old standing conjecture stated in the title.[2] "In general" means of course that the exceptional set is of the first category in one of the usual "natural" topologies (the strong neighborhood topology) for measure preserving transformations. The principal new and quite surprising fact used in the proof is that for any almost nowhere periodic measure preserving transformation T (and *a fortiori* for any mixing T) the set of all conjugates of T(i.e. the set of all STS^{-1}) is everywhere dense. It is this possibility of a dense conjugate class in a comparatively well behaved topological group (a rather natural generalization of the finite symmetric groups) that is contrary to naive intuition.

Let G be the group of all measure preserving transformations of the unit interval.[3] For any $S \epsilon G$, measurable set a, and positive number ϵ, write

$$N(S) = N(S; a, \epsilon) = \{T: | Sa - Ta | < \epsilon\}.^4$$

A unique topology (called the neighborhood topology) is defined in G by the requirement that the collection of all sets of the form $N(S)$ be a subbase for open sets.[5] (With this topology G becomes a complete topological group satisfying the second countability axiom.)[6] A permutation (or more precisely a dyadic permutation of rank m) is a measure preserving transformation which maps each dyadic interval of rank m onto itself or onto another one by a translation. (A dyadic interval of rank m is an interval of the form $(k/2^m, (k + 1)/2^m)$, $k = 0, 1, \cdots, 2^m - 1$, and a dyadic set of rank m is a union of such dyadic intervals.) A dyadic neighborhood is a set of the form $\{T: | Pa_i - Ta_i | < \epsilon, i = 1, \cdots, n\}$ where P is a permutation and the a_i are dyadic sets. The dyadic

---

[1] *Approximation theories for measure preserving transformations*, Transactions of the A. M. S., vol. 55 (1944), pp. 1–18, In the sequel I shall refer to this paper as (A).

[2] See Eberhard Hopf, *Ergodentheorie*, Berlin, 1937, p. 41. Cf. also (A), §1. The word 'mixing' is used in this paper in the weak sense of Hopf.

[3] It is known that the apparently great restriction to the unit interval is in fact very slight: measure theoretically most interesting spaces are isomorphic to each other and hence to the unit interval. See Paul R. Halmos and John von Neumann, *Operator methods in classical mechanics, II*, Annals of Math., vol. 43 (1942), Theorems 1 and 2.

[4] The symbol $\{—:—\}$ denotes the set of all those objects named before the colon which satisfy the condition stated after it. $|—|$ is used to denote the measure of a measurable set and $a - b$ standards for the symmetric difference of two sets, $a - b = ab' \cup a'b$, where a' is the complement of a.

[5] See Solomon Lefschetz, *Algebraic Topology*, New York, 1942, p. 6.

[6] See (A), Theorems 2 and 3.

neighborhoods form a base for the open sets of G. A measure preserving transformation T is almost nowhere periodic (or for brevity non periodic) if the set of those ξ, $0 \leqq \xi \leqq 1$, for which there is a positive integer n such that $T^n\xi = \xi$, has measure zero. Finally I define the symbols $d(S, T)$ and $\delta(S, T)$ by the formulae

$$d(S, T) = |\{\xi : S\xi \neq T\xi\}|, \quad \delta(S, T) = \sup \{ | Sa - Ta | \}.$$

In terms of these definitions it is now possible to state the following four results (proved in (A)) which will be needed in the sequel.

(1) *For every non periodic T and positive integer n there is a measurable set b such that $b, Tb, \cdots, T^{n-1}b$ are pairwise disjoint and $1/n \geqq |b| \geqq 1/(2n-1)$.*

(2) *Every dyadic neighborhood contains cyclic permutations of arbitrarily large ranks.*

(3) *If T has (almost everywhere) exactly the period m then there exists a measurable set b such that $b, Tb, \cdots, T^{m-1}b$ are pairwise disjoint and $|b| = 1/m$.* (See (A), p. 18).

(4) *Both $d(S, T)$ and $\delta(S, T)$ are group invariant metrics for G (giving rise, however, to a topology different from the neighborhood topology) and, for all S and T, $\delta(S, T) \leqq d(S, T)$.*

LEMMA 1. *For every non periodic T and positive integer n there exists a measure preserving transformation S and a measurable set a such that (i) $Sa = a$, (ii) S has exactly the period n in a, (iii) $|a| > 1/2$, (iv) $d(S, T) \leqq 2/n$, and (v) S is nowhere periodic in a'.*[7]

PROOF: Choose b in accordance with (1) so that $b, Tb, \cdots, T^{n-1}b$ are pairwise disjoint and $|b| \geqq 1/(2n-1)$. Write $a = b \cup Tb \cup \cdots \cup T^{n-1}b$; then $|a| \geqq n/(2n-1) > 1/2$. For $\xi \in b \cup Tb \cup \cdots \cup T^{n-2}b$ define $S\xi = T\xi$; for $\xi \in T^{n-1}b$ define $S\xi = T^{-n+1}\xi$. It follows that (*i*), (*ii*), and (*iii*) are satisfied: it remains to extend the definition of S in such a way as to satisfy (iv) and (v).

Consider $e = a' \cdot T^{-1}a'$ and $Te = a' \cdot Ta'$. Since $a' = e \cup (a' \cdot T^{-1}a)$ and $a' = Te \cup (a' \cdot Ta)$ it follows that $x = a' \cdot T^{-1}a$ and $y = a' \cdot Ta$ have the same measure. For $\xi \in e$ define $S\xi = T\xi$; it remains now to define S on x so that $Sx = y$. Before doing this I remark that (iv) is already achieved. This follows from the fact that

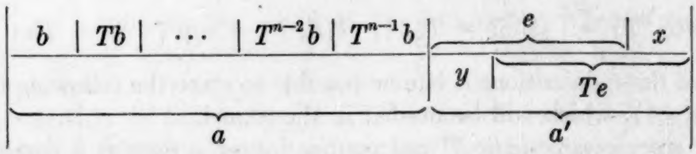
$$\{\xi : S\xi \neq T\xi\} \subset (T^{n-1}b) \cup (a' \cdot T^{-1}a),$$

whence (using the fact that T is measure preserving and that $a' \cdot T^{-1}a \subset T^{-1}b$)

$$d(S, T) \leqq 2|b| \leqq 2/n.$$

[7] This result is a strengthening of Lemma 3 of (A) and the first two paragraphs of the proof here given are a slight modification of the discussion in (A). A modification in this, as well as in several other results and definitions cited from (A), is necessary because the language and notation of (A) are considerably more algebraic than those of the present paper.

It may be assumed without loss of generality that x and y are disjoint, for otherwise S may be defined in xy to be any non periodic transformation and thus the problem reduces to finding a suitable transformation from xy' to $x'y$. This assumption makes it possible to draw a schematic sketch of all steps considered so far.



Let c be the set of those points $\xi \epsilon e$ for which $T^n\xi \epsilon e$ for all $n = 1, 2, \cdots$. Since $Tc \subset c$, the fact that T is measure preserving implies that (after possibly discarding a set of measure zero from c) $Tc = c$. It follows (since $c \subset e$) that $c \subset Te$. In other words almost every point of y gets out of e (and hence into x) in a finite number of steps. Write

$$y_k = \{\xi : \xi \epsilon y, T^i\xi \epsilon e \text{ for } i < k, T^k\xi \epsilon x\},$$

and define for $\xi \epsilon y_k$, $Q\xi = T^k\xi$. Let P be any non periodic transformation of y into itself and define for $\xi \epsilon x$, $S\xi = PQ^{-1}\xi$. S is now everywhere defined and it remains to prove that it is non periodic.

If $\xi \epsilon Te$ is such that $T^n\xi \epsilon Te$ for all $n > 0$ then (by an argument similar to that in the preceding paragraph) except for a set of ξ's of measure zero ξ and all its T transforms must belong to e also and consequently $S^n\xi = T^n\xi$. In this case the non periodicity of T ensures that of S. Hence it is sufficient to show that no $\xi \epsilon y$ has a finite orbit under S. To prove this take $\xi \epsilon y_k$. Then since $T^i\xi \epsilon e$ for $i = 0, \cdots, k-1$, $S^i\xi = T^i\xi$. Also since $T^i\xi \epsilon Te$ for $i = 1, \cdots, k$, $T\xi, \cdots, T^k\xi$ do not belong to y. Hence ξ is not only not periodic of some period $\leqq k$ but it does not even return to y in k steps. At the $(k+1)^{\text{th}}$ step (since $S^k\xi = T^k\xi = Q\xi \epsilon x$), $S^{k+1}\xi = PQ^{-1}T^k\xi$, whence, using the definition of Q, $S^{k+1}\xi = P\xi$. In other words: the first time that a point of y returns to y it is transformed by P. Hence the S-orbit of ξ contains the (infinite) P-orbit of ξ and cannot therefore be finite. This completes the proof of Lemma 1.

Lemma 2. *For every non periodic T and positive integer n there exists a measure preserving transformation S which has (almost) everywhere the period n and for which $d(S, T) \leqq 4/n$.*

Proof: The proof of this lemma is an application of Lemma 1 and consists of iterating that former result an infinite number of times. More precisely: after one application of Lemma 1 one obtains a set a_1 and a transformation S_1 with properties there described. The second time Lemma 1 is applied to S_1 on a_1' to find a set $a_2 \subset a_1'$ and a transformation S_2 such that S_2 has the period n on a_2, $|a_2| \geqq (1/2)|a_1'|$, and $d(S_1, S_2) \leqq (2/n)|a_1'|$. Proceeding in this way by induction one obtains a sequence of disjoint measurable sets $a_1, a_2, \cdots$ and a sequence of measure preserving transformations $S_1, S_2, \cdots$ such that

$d(S_i, S_{i+1}) \leqq (2/n) \mid (a_1 \cup \cdots \cup a_i)' \mid$, $\mid a_{i+1} \mid \geqq \mid (a_1 \cup \cdots \cup a_i)' \mid /2$, $S_i a_i = a_i$, and S_i has period n in a_i. It follows also (by induction) that $\mid a_1 \cup \cdots \cup a_i \mid > 1/2 + \cdots + 1/2^i = 1 - 1/2^i$. Hence the union of all the a's is (except possibly for a set of measure zero) the entire unit interval and therefore a measure preserving transformation S of the entire interval may be defined by writing, for $\xi \in a_i$, $S\xi = S_i\xi$. S will have almost everywhere the period n and, moreover,

$$d(T, S) \leqq d(T, S_1) + d(S_1, S_2) + d(S_2, S_3) + \cdots$$
$$\leqq (2/n)(1 + 1/2 + 1/2^2 + 1/2^3 + \cdots) = 4/n.$$

With these preliminary results out of the way I am now in a position to state and prove the first main theorem of this paper.

THEOREM 1. *The conjugate class of any non periodic measure preserving transformation $T_0 \in G$ is (in the neighborhood topology) everywhere dense in G.*

PROOF: Let $N = N(P) = \{T: \mid Pa_i - Ta_i \mid < \epsilon, i = 1, \cdots, n\}$ be any dyadic neighborhood of some permutation P: it is to be proved that for some $R \in G$, $RT_0R^{-1} \in N$. Write $M = \{T: \mid Pa_i - Ta_i \mid < \epsilon/2, i = 1, \cdots, n\}$ and apply to M the result (2) quoted from (A). It follows that M contains a cyclic permutation Q of rank m greater than the ranks of all a_i and such that $1/2^{m-3} < \epsilon$. Then Lemma 2 may be applied to T_0 to obtain an S which has (almost) everywhere the period 2^m and for which $d(S, T_0) \leqq 4/2^m < \epsilon/2$.

Observe next that Q and S are conjugate in G. For denote by $q_0, \cdots, q_{r-1}$ ($r = 2^m$) the dyadic intervals of rank m, so arranged that $Qq_i = q_{i+1}$ ($i \bmod r$) and denote by s_0 a set of measure $1/2^m$ which is disjoint from S^is_0 for $i = 1, \cdots, r - 1$. (The existence of such an s_0 is guaranteed by (3)). Write $s_i = S^is_0$, $i = 1, \cdots, r - 1$, and let R be any (measure preserving) transformation taking s_0 into q_0. To extend the definition of R write, for any $\xi \in s_i$, $R\xi = Q^iRS^{-i}\xi$. Schematically:

$$\begin{tikzcd}
s_0 \arrow[r, "S"] \arrow[d, "R"'] & s_1 \arrow[r, "S"] \arrow[d, "R"'] & s_2 \cdots s_{r-2} \arrow[r, "S"] \arrow[d, "R"'] & s_{r-1} \arrow[d, "R"'] \\
q_0 \arrow[r, "Q"'] & q_1 \arrow[r, "Q"'] & q_2 \cdots q_{r-2} \arrow[r, "Q"'] & q_{r-1}
\end{tikzcd}$$

It is quite easy to see that $Q = RSR^{-1}$.

The proof can now be quickly finished. Since $d(S, T_0) < \epsilon/2$ and since (by (4)) d is invariant under the group operations,

$$d(Q, RT_0R^{-1}) = d(RSR^{-1}, RT_0R^{-1}) = d(S, T_0) < \epsilon/2.$$

Using the definition of δ and the inequality stated in (4) it follows that

$$\mid Pa_i - RT_0R^{-1}a_i \mid \leqq \mid Pa_i - Qa_i \mid + \mid Qa_i - RT_0R^{-1}a_i \mid$$
$$\leqq \mid Pa_i - Qa_i \mid + \delta(Q, RT_0R^{-1})$$
$$< \epsilon/2 + d(Q, RT_0R^{-1}) < \epsilon.$$

Consequently $RT_0R^{-1} \in N(P)$; q.e.d.

To establish the residual character of any set M of measure preserving transformations it is sufficient, in view of the theorem just proved, to show that (i) M contains a non periodic transformation, (ii) M is self conjugate, and (iii) M is a G_δ.[8] In the sequel I shall show that the set M of mixing transformations has these three properties.

Recall first the definition of mixing. A measure preserving transformation is mixing if there exists a set I_0 of positive integers of density zero such that for any two measurable sets a and b

$$\lim_{n \notin I_0} | T^n a \cdot b | = | a | \cdot | b |.^{9}$$

The properties (i) and (ii) of M are easy consequences of this definition. (i) Not only does M contain non periodic transformations but in fact every mixing T is non periodic.[10] For if the set of those ξ whose period is say m had positive measure then (3) (applied to T considered on this set only) would yield a set c of positive measure with the properties $T^m c = c$, $T^j c \cdot c = 0$ for $0 < j < m$. Writing $a = b = c$, it follows that $| T^n a \cdot b | = | c |$ or 0 according as $n \equiv 0 \pmod{m}$ or not. Since this sort of behavior clearly excludes the possibility of the limiting property of mixing transformations it follows that a T which is periodic on a set of positive measure cannot be mixing. (ii) If T is mixing and $S = RTR^{-1}$ then S is mixing. For

$$\begin{aligned} | S^n a \cdot b | &= | (RTR^{-1})^n a \cdot b | = | RT^n R^{-1} a \cdot b | \\ &= | R(T^n R^{-1} a \cdot R^{-1} b) | = | T^n(R^{-1}a) \cdot (R^{-1}b) |; \end{aligned}$$

the limiting property of T (with a, b replaced by $R^{-1}a$, $R^{-1}b$) together with the fact that R is measure preserving yields the desired conclusion.

To prove (iii), i.e. to prove that M is a G_δ, it is necessary to use the known Hilbert space characterization of mixing. A measure preserving transformation T may be regarded as a unitary operator on the complex Hilbert space L_2, by defining $Tf(\xi) = f(T\xi)$. (The unitary character of T follows from the fact that measure and, consequently, integral are invariant.) Mixing in Hilbert space terms amounts to the existence of a set I_0 of positive integers of density zero such that for any two functions $f, g \in L_2$,

$$\lim_{n \notin I_0} (T^n f, g) = (f, 1)(1, g)$$

(where (f, g) is the inner product $\int_0^1 f(\xi)\overline{g(\xi)}\, d\xi$ and 1 stands for the function identically equal to 1, so that $(f, 1) = \int_0^1 f(\xi)\, d\xi$ and $(1, g) = \int_0^1 \overline{g(\xi)}\, d\xi$.)

---

[8] Cf. the proof of Theorem 6, (A).

[9] See Hopf, op. cit., p. 36. For all concepts and results quoted in this paragraph and the next the reader is referred to Chapter III of Hopf's book.

[10] To see that this statement is really stronger than the original formulation of (i) it is necessary to make the non trivial observation that there exists at least one mixing transformation. See for example Paul R. Halmos, *On automorphisms of compact groups*, Bulletin of the A. M. S., vol. 49 (1943), pp. 619–624.

From the point of view of the preceding paragraph the set G of all measure preserving transformations becomes a subset of the set U of all unitary operators. On this latter set von Neumann's "strong neighborhood topology" may be defined: a subbasic neighborhood of $T_0 \epsilon U$ is of the form

$$N(T_0) = N(T_0\,; f, \epsilon) = \{T : \| T_0 f - Tf \| < \epsilon\},$$

where $f \epsilon L_2$, $\epsilon > 0$, and $\| f \|$ stands, as usual, for $(f, f)^{1/2}$. I shall need to make use of the fact that this topology when specialized to the subset G of U coincides with the neighborhood topology defined earlier. (This fact is proved by showing that a topology equivalent to the "strong neighborhood topology" of von Neumann is obtained if f is restricted to be the characteristic function of a measurable set.[11]) Two further facts will be used in the proof that follows: first that for fixed f, $g \epsilon L_2$ the function $\varphi(T) = T(f, g)$ is a continuous function of T (in the neighborhood topology) and, last but far from least, that T is mixing if and only if it has no non trivial proper values, i.e. $Tf = \lambda f$ $(f \neq 0)$ implies that f is (almost everywhere) equal to the constant $\lambda = 1$. (This last fact is known as the mixing theorem.[12])

The remainder of the proof (that M is a G_δ) is somewhat analogous to the proof of Theorem 6 of (A). Let $f_1, f_2, \cdots$ be a dense sequence in L_2; write

$$E(i, j, m, n) = \{T : | (T^n f_i, f_j) - (f_i, 1)(1, f_j) | < 1/m\}$$

and

$$F = \bigcap_i \bigcap_j \bigcap_m \bigcup_n E(i, j, m, n).$$

The fact that (Tf, g) is a continuous function of T implies that $E(i, j, m, n)$ is open and therefore that F is a G_δ. The limiting property of mixing T's shows that $M \subset F$: I shall now show that, conversely, if T is not mixing then T is not in F.

If T is not mixing then, by the mixing theorem, there is an $f \neq 0$ and a complex number λ of modulus one (since T is unitary) such that $Tf = \lambda f$. Without loss of generality it may be assumed that f is orthogonal to the trivial proper function 1, $(f, 1) = 0$, and that f is normalized, $\| f \| = 1$. Choose now an i for which $\| f - f_i \| < .1$ and choose $j = i$, $m = 2$. I shall show that T does not belong to F. Observe first that $\| f_i \| \leqq \| f \| + \| f_i - f \| \leqq 1.1$. It follows that

$$\begin{aligned} 1 &= | (T^n f, f) - (f, 1)(1, f) | \\ &\leqq | (T^n f, f) - (T^n f, f_i) | + | (T^n f, f_i) - (T^n f_i, f_i) | + | (T^n f_i, f_i) - (f_i, 1)(1, f_i) | \\ &\qquad + | (f_i, 1)(1, f_i) - (f_i, 1)(1, f) | + | (f_i, 1)(1, f) - (f, 1)(1, f) | \end{aligned}$$

---

[11] See John von Neumann, *Zur Algebra der Funktionaloperationen und Theorie der normalen Operatoren*, Math. Ann., vol. 102 (1929), p. 386.

[12] Hopf, op. cit., p. 37.

$$\leqq \| f - f_i \| + \| f - f_i \| \| f_i \| + \| f_i \| \| f_i - f \| + \| f_i - f \| + | (T^n f_i, f_i) - (f_i, 1)(1, f_i) |$$

$$\leqq .1 + .11 + .11 + .1 + | (T^n f_i, f_i) - (f_i, 1)(1, f_i) |$$

$$< .5 + | (T^n f_i, f_i) - (f_i, 1)(1, f_i) |.$$

In other words

$$| (T^n f_i, f_i) - (f_i, 1)(1, f_i) | > .5 = 1/m$$

for all n. Hence:

THEOREM 2. *In the neighborhood topology of G the set M of mixing transformations is a residual G_δ.*

SYRACUSE UNIVERSITY

ANNALS OF MATHEMATICS
Vol. 45, No. 4, October, 1944

MEASURING THE SURFACE AREA OF A CONVEX BODY

BY P. A. P. MORAN

(Received March 20, 1944)

In applied sciences it is sometimes desirable to measure the surface area of a small body. It is difficult to do this directly, but if the body is convex we may use a well-known theorem on convex bodies. If K is such a body in three-dimensional space, let S be its surface area, $A(n)$ the area of its projection on a plane perpendicular to the vector n, and A the mean of $A(n)$ for all directions of n. Then the theorem states

$$S = 4A.$$

To obtain a value for this mean in practice, we may take the average of the projections in a finite number of directions, and the question then arises as to the error which may result in such a process. In what follows we suppose the projected area to be measured in a set of directions which is in one case the set of directions of the normals to the faces of a dodecahedron, and in a second case the set of directions of the normals to the faces of an icosahedron. We shall then find exact upper and lower bounds for the ratio between the estimated area and the exact area. In practice the area of the projection might be measured by measuring the shadow of the body when placed in the path of a suitably directed beam of light, or by observing the body through a suitably directed optical instrument in whose eyepiece is placed a graticule ruled in squares. For a bibliography on the experimental side, see Markwick.[1]

H. Steinhaus[2] has considered an analogous problem concerning a convex figure in a plane. If F is such a figure and $f(\theta)$ is the length of its projection on a line making an angle θ with a fixed direction, it is known that the length of the perimeter of F is π times the mean of $f(\theta)$ for all θ. Steinhaus replaces this mean by the average of the values of $f(\theta)$ for six different values separated by intervals of 30°, and then finds exact upper and lower bounds for the ratio between the estimated length and the true length. His method is trigonometrical, and does not extend readily to the three-dimensional case. We employ a vectorial method, and the ideas of this method enable us to extend readily Steinhaus's result to the case of an arbitrary number of equally spaced directions in a plane.

We begin by considering the dodecahedral case. Let T be an element of area whose normal is a unit vector $\mathbf{l}$. We suppose T to be of unit area and to be placed at the centre of a dodecahedron so that its projection in the direction of the normals to the twelve faces can be considered. Let these faces be numbered from 1 to 12. Let the line of the vector $\mathbf{l}$ intersect the face 1, and suppose the

[1] A. H. D. Markwick, *Chemistry and Industry*, XV, 206 (1937).
[2] H. Steinhaus, *Akad. d. wiss. Leipzig, Ber.*, 82, 120–130 (1930).

faces adjoint to this face be 2, 3, $\cdots$ 6 (Fig. 1). Let the faces opposite to 1, 2, $\cdots$ 6 be 7, 8, $\cdots$ 12 respectively. Let $\mathbf{l}_1, \cdots \mathbf{l}_{12}$ be unit vectors directed from the centre of the dodecahedron in the directions of the normals to the faces 1, 2, $\cdots$ 12. Then the mean of the areas of the projections of the element T in the twelve directions is given by

$$\frac{1}{12}\sum_{i=1}^{12} |\mathbf{l}\cdot\mathbf{l}_i|$$

since each term in the sum is the area of the projection in the direction of the corresponding vector. We now find upper and lower bounds for this sum. Let the face 1 have vertices A, B, C, D, E, where CD, DE, EA, AB and BC are edges in common with the faces 2, 3, 4, 5 and 6 respectively, and let the middle points of CD, DE, EA, AB and BC be A', B', C', D', E'. Then two cases arise.

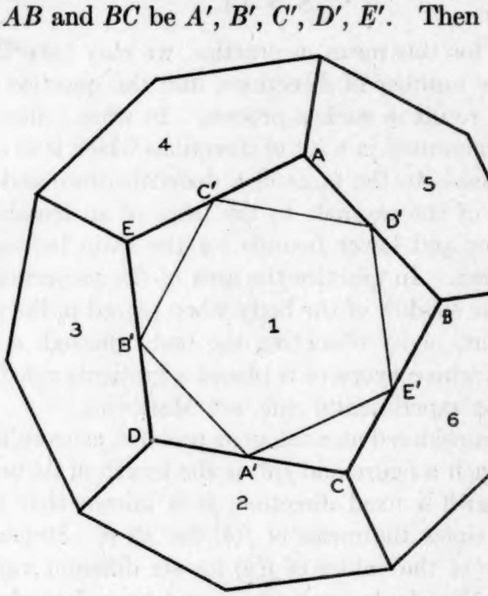


Fig. 1

I. The line of the vector $\mathbf{l}$ intersects the dodecahedron in a point inside the pentagon $A'B'C'D'E'$.

If this is true the vectors $\mathbf{l}_1, \cdots \mathbf{l}_6$ all lie on one side of the plane of the element of unit area, and the vectors $\mathbf{l}_7, \cdots \mathbf{l}_{12}$ on the other. By symmetry we can clearly confine ourselves to those on the same side as the vector $\mathbf{l}$. Then the mean of the projected areas is

$$\frac{1}{6}\sum_{i=1}^{6} |\mathbf{l}\cdot\mathbf{l}_i| = \frac{1}{6}\sum_{i=1}^{6} \mathbf{l}\cdot\mathbf{l}_i$$

$$= \frac{1}{6}\mathbf{l}\cdot\left\{\sum_{i=1}^{6} \mathbf{l}_i\right\}.$$

Now the vector $\sum_{i=1}^{6} \mathbf{l}_i$ is a vector in the direction of $\mathbf{l}_1$, and has magnitude $1 + 5 \cos \theta$, where θ is the angle between the vectors associated with adjacent faces of the dodecahedron. A little calculation shows this angle to be 63.435°, and so $\cos \theta = .4472$. Then

$$\frac{1}{6} \sum_{i=1}^{6} | \mathbf{l} \cdot \mathbf{l}_i | = \tfrac{1}{6} \cos \epsilon (1 + 5 \cos \theta)$$

where ϵ is the angle between the vectors $\mathbf{l}$ and $\mathbf{l}_1$. The largest attainable value of this expression is

$$K_1 = \tfrac{1}{6}(1 + 5 \cos \theta)$$
$$= .5393$$

and the smallest is

$$K_2 = \tfrac{1}{6} \cos (\theta/2)(1 + 5 \cos \theta)$$
$$= .4588.$$

The area of one side of the element of area will be twice the mean projected area, and so if S_1 is the estimated area and S the true area, we have

$$.918 = 2K_2 \leqq \frac{S_i}{S} \leqq 2K_1 = 1.079.$$

Moreover these limits are clearly the best possible.

II. The line of the vector $\mathbf{l}$ intersects the dodecahedron in a point outside the pentagon $A'B'C'D'E'$ and inside the pentagon $ABCDE$.

We can suppose by symmetry that the point of intersection with the dodecahedron is within the triangle $CA'E'$. Then the vectors $\mathbf{l}_1$, $\mathbf{l}_2$, $\mathbf{l}_3$, $\mathbf{l}_5$, $\mathbf{l}_6$ and $\mathbf{l}_{10}$ lie on one side of the plane of the element of area. The mean of the projections in all twelve directions is

$$\frac{1}{12} \sum_{i=1}^{12} | \mathbf{l} \cdot \mathbf{l}_i | = \tfrac{1}{6}\{| \mathbf{l} \cdot \mathbf{l}_1 | + | \mathbf{l} \cdot \mathbf{l}_2 | + | \mathbf{l} \cdot \mathbf{l}_3 | + | \mathbf{l} \cdot \mathbf{l}_5 | + | \mathbf{l} \cdot \mathbf{l}_6 | + | \mathbf{l} \cdot \mathbf{l}_{10} |\}$$
$$= \tfrac{1}{6}\mathbf{l} \cdot \{\mathbf{l}_1 + \mathbf{l}_2 + \mathbf{l}_3 + \mathbf{l}_5 + \mathbf{l}_6 + \mathbf{l}_{10}\}$$
$$= \tfrac{1}{6}\mathbf{l} \cdot \mathbf{L}$$

where $\mathbf{L}$ is a vector which is the sum of the vectors in the bracket and lies in the direction from the centre of the dodecahedron to the point C. Now the minimum of the above expression occurs when the vector $\mathbf{l}$ points at A' or E', i.e. in the same direction as gave the minimum in case 1, and therefore giving the same result. We need therefore only calculate the maximum value of $\tfrac{1}{6}\mathbf{l} \cdot \mathbf{L}$, which is $\tfrac{1}{6} | \mathbf{L} |$ and

$$| \mathbf{L} | = 3(\cos \phi_1 + \cos \phi_2)$$

where ϕ_1 is the angle between $\mathbf{l}_1$ and $\mathbf{L}$, and ϕ_2 is the angle between $\mathbf{l}_3$ and $\mathbf{L}$.

A little calculation shows that

$$\tfrac{1}{3}\,|\,\mathbf{L}\,| = .984$$

and we have in this case

$$\frac{S_1}{S} \leqq .984 < 1.079$$

and therefore in whatever direction the vector normal to the element lies we find that the inequality of case I holds. The surface of any convex body can be considered as made up of infinitesimal elements, and so the above inequality holds for any convex body and the result is the best possible, since the bounds can actually be obtained when the convex body is flat.

We now consider the similar problem where we measure the projections in the directions of the normals to the faces of an icosahedron. Let the unit vectors in the direction of the faces be $\mathbf{l}_1, \cdots \mathbf{l}_{20}$. Suppose once again we have an element of unit area at the centre of the icosahedron and that the unit vector in the direction of the normal to the element of unit area is $\mathbf{l}$. Let the line of the vector $\mathbf{l}$ intersect the face ABC, and suppose this face is numbered 1 corresponding to the vector $\mathbf{l}_1$. Let the faces having BC, CA and AB in common with ABC be numbered 2, 3, 4 respectively, and similarly for the corresponding vectors (Fig. 2). Let the other vectors associated with the faces having angles at the vertices A, B and C be $\mathbf{l}_5$ and $\mathbf{l}_6$, $\mathbf{l}_7$ and $\mathbf{l}_8$, $\mathbf{l}_9$ and $\mathbf{l}_{10}$ respectively. Let the vectors associated with the faces opposite to the faces 1, $\cdots$ 10 be $\mathbf{l}_{11}, \cdots \mathbf{l}_{20}$ respectively. Let A', B', C' be the centres of the edges BC, CA and AB, and suppose $A'S$ is a line on the face ABC such that it lies in a plane through the centre of the icosahedron perpendicular to the vector $\mathbf{l}_5$. Similarly, let $A'P$, $B'M$, $B'R$, $C'Q$, $C'L$ be lines lying in planes perpendicular to $\mathbf{l}_6, \mathbf{l}_7, \mathbf{l}_8, \mathbf{l}_9, \mathbf{l}_{10}$. Let $B'R$ and $C'Q$ intersect in D, $A'P$ and $B'M$ in E, and $C'L$ and $A'S$ in F. Then $DC'FA'EB'$ is an irregular hexagon. By symmetry it is sufficient to consider the following cases—

1. The line of the vector $\mathbf{l}$ intersects the face ABC within the irregular hexagon $DC'FA'EB'$. Then the vectors $\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3, \mathbf{l}_4, \mathbf{l}_5, \mathbf{l}_6, \mathbf{l}_7, \mathbf{l}_8, \mathbf{l}_9, \mathbf{l}_{10}$ all lie on one side of the plane of the element of area and the sum

$$\sum_{i=1}^{20} |\,\mathbf{l}\cdot\mathbf{l}_i\,| = 2\sum_{i=1}^{10} |\,\mathbf{l}\cdot\mathbf{l}_i\,| = 2\mathbf{l}\cdot\left(\sum_{i=1}^{10} \mathbf{l}_i\right) = 2\mathbf{l}\cdot\mathbf{L}_1$$

where $\mathbf{L}_1$ is a vector from the centre of the icosahedron in the direction $\mathbf{l}_1$, and having the magnitude

$$1 + 3\cos\alpha + 6\cos\beta$$

where α is the angle between the vectors $\mathbf{l}_1$ and $\mathbf{l}_2$, and β is the angle between the vectors $\mathbf{l}_1$ and $\mathbf{l}_5$. Calculation shows that $\cos\alpha = .7465$ and $\cos\beta = .3333$. Then

$$2\,\mathbf{l}\cdot\mathbf{L} = 2\cos\epsilon(1 + 3\cos\alpha + 6\cos\beta)$$

where ϵ is the angle between **l** and **L**. The maximum value of this product is

$$2(1 + 3 \cos \alpha + 6 \cos \beta)$$

and the minimum value

$$2 \cos \theta_1 (1 + 3 \cos \alpha + 6 \cos \beta)$$

where θ_1 is the angle between $\mathbf{l}_1$ and the line from the centre of the icosahedron to the point D (or E, or F) for it is easy to see that the point D is further from the centre of the triangle than A', B' or C'. Then if the conditions of case I

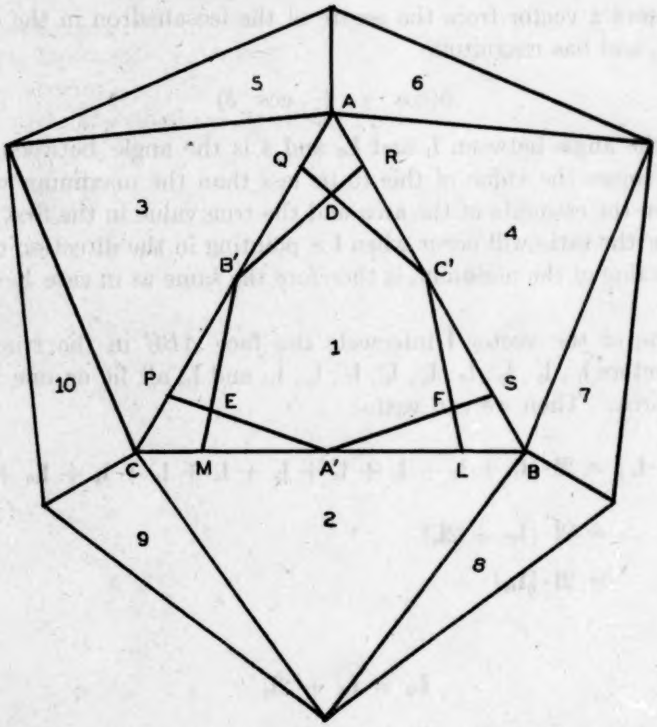


Fig. 2

are satisfied the ratio of the estimated to the true value of the element of area lies between

$$\tfrac{1}{5}(1 + 3 \cos \alpha + 6 \cos \beta)$$

and

$$\tfrac{1}{5} \cos \theta_1 (1 + 3 \cos \alpha + 6 \cos \beta)$$

and some calculation shows these to be 1.048 and .957 respectively.

2. The line of the vector **l** intersects the face ABC in the lozenge shaped region $ARDQ$. Then the vectors $\mathbf{l}_1$, $\mathbf{l}_2$, $\mathbf{l}_3$, $\mathbf{l}_4$, $\mathbf{l}_5$, $\mathbf{l}_6$, $\mathbf{l}_7$, $\mathbf{l}_{10}$, $\mathbf{l}_{18}$ and $\mathbf{l}_{19}$ lie on

one side of the element of area and the remainder on the other side, and we can write

$$\sum_{i=1}^{20} |\mathbf{l}\cdot\mathbf{l}_i| = 2\mathbf{l}\cdot\{\mathbf{l}_1 + \mathbf{l}_2 + \mathbf{l}_3 + \mathbf{l}_4 + \mathbf{l}_5 + \mathbf{l}_6 + \mathbf{l}_7 + \mathbf{l}_{10} + \mathbf{l}_{18} + \mathbf{l}_{19}\}$$

$$= 2\mathbf{l}\cdot\mathbf{L}_2$$

where

$$\mathbf{L}_2 = \mathbf{l}_1 + \mathbf{l}_2 + \mathbf{l}_3 + \mathbf{l}_4 + \mathbf{l}_5 + \mathbf{l}_6 + \mathbf{l}_7 + \mathbf{l}_{10} + \mathbf{l}_{18} + \mathbf{l}_{19}$$

and is therefore a vector from the centre of the icosahedron in the direction of the point A, and has magnitude

$$5(\cos\ \gamma + \cos\ \delta)$$

where γ is the angle between $\mathbf{l}_1$ and $\mathbf{L}_2$ and δ is the angle between $\mathbf{l}_2$ and $\mathbf{L}_2$. Calculation shows the value of this to be less than the maximum value of the ratio between the estimate of the area and the true value in the first case. The minimum for the ratio will occur when $\mathbf{l}$ is pointing in the direction of the point D, and the value of the minimum is therefore the same as in case I.

3. The line of the vector $\mathbf{l}$ intersects the face ABC in the triangle RDC'. Then the vectors $\mathbf{l}_1$, $\mathbf{l}_2$, $\mathbf{l}_3$, $\mathbf{l}_4$, $\mathbf{l}_5$, $\mathbf{l}_6$, $\mathbf{l}_7$, $\mathbf{l}_8$, $\mathbf{l}_{10}$ and $\mathbf{l}_{19}$ all lie on one side of the element of area. Then we can write

$$\sum_{i=1}^{20} |\mathbf{l}\cdot\mathbf{l}_i| = 2\mathbf{l}\cdot\{\mathbf{l}_1 + \mathbf{l}_2 + \mathbf{l}_3 + \mathbf{l}_4 + \mathbf{l}_5 + \mathbf{l}_6 + \mathbf{l}_7 + \mathbf{l}_8 + \mathbf{l}_{10} + \mathbf{l}_{19}\}$$

$$= 2\mathbf{l}\cdot\{\mathbf{L}_2 + 2\mathbf{l}_8\}$$

$$= 2\mathbf{l}\cdot\{\mathbf{L}_3\}$$

where

$$\mathbf{L}_3 = \mathbf{L}_2 + 2\mathbf{l}_8.$$

A little calculation shows the absolute value of $\mathbf{L}_3$ to be less than that of $\mathbf{L}_1$, and so the maximum value of the ratio of the estimated to the true area is less in this case than in case I. Moreover the minimum will occur when the vector $\mathbf{l}$ points in the direction of some point on RD or DC' and so will not be smaller than the minimum in case I or II. Consequently in any case the ratio of the estimated to the true area will lie between 1.048 and .957. By summing all the elements of the surface of a convex body the same result is seen to hold for the latter, and this completes the discussion.

Similar methods can be applied to extend the results of Steinhaus for convex figures in a plane. Suppose that we measure the projection of such a body in n evenly spaced directions separated from one another by angles equal to $2\pi/n$. Two cases arise according as to whether n is even or odd. If n is odd the result

is the same as measuring the body in $2n$ evenly spaced directions. We therefore need only consider the latter case, and by applying an argument similar to the above we show very simply that the ratio of the estimated perimeter to the true perimeter must be between $\pi/(2n \sin \pi/2n)$ and $(\pi \cos \pi/2n)/(2n \sin \pi/2n)$.

For $n = 6$, these limits are 1.012 and .977
for $n = 10$, 1.004 and .992
for $n = 20$, 1.001 and .998.

These results also have applications to practical problems.[3]

ST. JOHN'S COLLEGE
CAMBRIDGE, ENGLAND

[3] H. König, *Arch. Eisenhüttenwes*, 7, 441–444 (1934).

YALE UNIVERSITY
JAN 27 1944
LIBRARY

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

SOLOMON LEFSCHETZ JOHN VON NEUMANN

FRÉDÉRIC BOHNENBLUST

WITH THE COÖPERATION OF THE

DEPARTMENT OF MATHEMATICS OF PRINCETON UNIVERSITY

AND

THE SCHOOL OF MATHEMATICS OF THE INSTITUTE FOR ADVANCED STUDY

AND

A. A. ALBERT, EMIL ARTIN, GARRETT BIRKHOFF, M. R. HESTENES[1], T. H. HILDEBRANDT[2], NATHAN JACOBSON, N. E. STEENROD, M. H. STONE, ANNA PELL-WHEELER[1], HASSLER WHITNEY[2], G. T. WHYBURN[2], OSCAR ZARISKI

([1] Representing the MATHEMATICAL ASSOCIATION OF AMERICA)
([2] Representing the AMERICAN MATHEMATICAL SOCIETY)

SECOND SERIES, VOL. 45, NO. 1
JANUARY, 1944

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE-2, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

COPYRIGHT, 1944, BY PRINCETON UNIVERSITY PRESS

ANNALS OF MATHEMATICS

Subscriptions: The subscription price is $5.00 a volume (four numbers); the price of a single copy is $1.50. An arrangement has been made with the MATHEMATICAL ASSOCIATION OF AMERICA whereby an individual member of the ASSOCIATION may subscribe to the ANNALS OF MATHEMATICS at one-half the regular price. Subscription orders, orders for back numbers, notices of change of address, and other business correspondence should be sent to the Princeton University Press, Princeton, N. J.

Editorial Correspondence: Editorial correspondence and manuscripts should be addressed to the ANNALS OF MATHEMATICS, Fine Hall, Princeton, N. J., U. S. A.

Preparation of Manuscripts: Manuscripts submitted to the ANNALS OF MATHEMATICS must be typewritten, double spaced, with generous margins. Figures and diagrams should be drawn in India ink on separate sheets of paper and of twice the size they are to be printed. The lettering is to be omitted from these drawings, but a sketch with complete lettering should be appended.

Footnotes should be numbered consecutively and typed, double spaced, on a sheet by themselves at the end of the manuscript—they will be inserted at the designated places in the text after galley proofs have been corrected. Footnotes should be reduced to a minimum and, whenever possible, replaced by references to a bibliography placed at the end of the paper. Formulae are undesirable in footnotes.

Authors are requested to keep in mind the typographical difficulties involved in setting complicated mathematical formulae. These mechanical difficulties may frequently be eliminated or reduced by the use of alternative notations, *e. g.* "exp" may be used to avoid a complicated exponent, and fractional exponents may be used to replace the troublesome radical sign. The use of unusual symbols is to be avoided. Symbols set one on top of another present a typographical difficulty which usually can be easily avoided by a stepwise arrangement, *e. g.* replace $\sum\limits_{1}^{n}$ by $\sum_1^n$, etc.

Proofs: Authors ordinarily receive only galley proofs. Stylistic changes in proofs must be avoided; authors will be charged for excessive corrections.

Reprints: Authors receive gratis, postage prepaid, one hundred reprints of each article; joint authors each receive fifty reprints.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Made in the United States of America

ANNALS OF MATHEMATICS STUDIES

1. **Algebraic Theory of Numbers**
   By HERMANN WEYL 227 *pp.* $2.35
2. **Convergence and Uniformity in Topology**
   By JOHN W. TUKEY 95 *pp.* $1.50
3. **The Consistency of the Continuum Hypothesis**
   By KURT GÖDEL 68 *pp.* $1.25
4. **An Introduction to Linear Transformations in Hilbert Space**
   By F. J. MURRAY 135 *pp.* $1.75
5. **The Two-Valued Iterative Systems of Mathematical Logic**
   By EMIL L. POST 122 *pp.* $1.75
6. **The Calculi of Lambda-Conversion**
   By ALONZO CHURCH 77 *pp.* $1.25
7. **Finite Dimensional Vector Spaces**
   By PAUL R. HALMOS, 201 *pp.* $2.35
8. **Metric Methods in Finsler Spaces and in the Foundations of Geometry**
   By HERBERT BUSEMANN 247 *pp.* $3.00
9. **Degree of Approximation by Polynomials in the Complex Domain**
   By W. E. SEWELL 246 *pp.* $3.00
10. **Topics in Topology**
    By SOLOMON LEFSCHETZ 139 *pp.* $2.00
11. **Introduction to Non-Linear Mechanics**
    By N. KRYLOFF and N. BOGOLIUBOFF 108 *pp.* $1.65
12. **Meromorphic Functions and Analytic Curves**
    By HERMANN WEYL 277 *pp.* $3.50

The prices above are subject to an educational discount of fifteen per cent.

PRINCETON UNIVERSITY PRESS
Princeton, New Jersey

PRINCETON MATHEMATICAL SERIES

Edited by

MARSTON MORSE H. P. ROBERTSON A. W. TUCKER

1. **The Classical Groups.**
   By HERMANN WEYL 302 *pages*, **$4.00**
2. **Topological Groups.**
   By L. PONTRJAGIN (translated by E. Lehmer) 299 *pages*, **$4.00**
3. **An Introduction to Differential Geometry.**
   By LUTHER PFAHLER EISENHART 304 *pages*, **$3.50**
4. **Dimension Theory.**
   By WITOLD HUREWICZ & HENRY WALLMAN 165 *pages*, **$3.00**
5. **Analytical Foundations of Celestial Mechanics.**
   By AUREL WINTNER 448 *pages*, **$6.00**
6. **Laplace Transform.**
   By DAVID VERNON WIDDER 406 *pages*, **$6.00**
7. **Integration**
   By E. J. MCSHANE In press

The prices above are subject to an educational discount of fifteen per cent.

PRINCETON UNIVERSITY PRESS
Princeton, New Jersey

THE JOURNAL *of* SYMBOLIC LOGIC

Edited by Alonzo Church and Ernest Nagel
Managing Editor, Helen C. Brodie
Consulting Editors: Paul Bernays, C. A. Baylis, Evert Beth, C. G. Hempel, C. H. Langford, Everett J. Nelson, Max Black, G. D. W. Berry, László Kalmár, Rózsa Péter, H. E. Vaughan, F. B. Fitch, J. C. C. McKinsey

An international journal, publishing contributions to symbolic logic in English, French, and German. Volumes I and III together include a complete bibliography of symbolic logic for the period 1666–1935, indexed by authors and by subjects. A complete current bibliography of literature in the field, both books and articles, from Jan. 1, 1936, is provided by prompt publication of critical reviews.

Published Quarterly by the

ASSOCIATION FOR SYMBOLIC LOGIC

Annual membership, including subscription to the Journal, $3.00. Current subscription, $3.00. Completed volumes, $4.00. Part I of the Bibliography, $1.50; on rag paper $2.00. Part II of the Bibliography, $1.00; on rag paper, $1.25.

Application for membership or subscription should be sent to Helen C. Brodie, Secretary-Treasurer, 29 Washington Square West, New York, 11, N. Y.

Copies of the following memoirs can be obtained by addressing

for the United States, Canada, and South America, The Nordemann Publishing Company, Inc., 215 Fourth Ave., New York, N. Y.

An introduction to the theory of elliptic functions. By GOSTA MITTAG-LEFFLER. Authorized translation by EINAR HILLE. 81 pages. Price 90 cents.

On various conceptions of correlation. By F. M. WEIDA. 37 pages. Price 75 cents.

Contact transformations. By L. P. EISENHART. 40 pages. Price 80 cents.

A survey of the theory of small samples. By P. R. RIDER. 52 pages. Price $1.00.

Tauberian theorems. By NORBERT WIENER. 100 pages. Price $1.00.

Untersuchungen über Gestalt und Lage abgeschlossener Mengen beliebiger Dimension. By PAUL ALEXANDROFF. 86 pages. Price $1.00.

Plateau's problem and Dirichlet's principle. By RICHARD COURANT. 46 pages. Price 90 cents.

Minimal surfaces of higher topological structure. By J. DOUGLAS. 94 pages. Price $1.00.

Sur quelques notions fondamentales dans la théorie générale des operations linéaires. By FRÉDÉRIC RIESZ. 33 pages. Price 70 cents.

La théorie du corps de classes. By C. CHEVALLEY. 25 pages. Price 50 cents.

MATHEMATICAL REVIEWS

offers

prompt review of the mathematical literature of the world
complete subject and author indices in last issue of volume
microfilm or photoprints of most articles reviewed (at cost)

at the very low rate of $13 per year

Publication of this journal is sponsored by the American Mathematical Society, Mathematical Association of America, London Mathematical Society, Academia Nacional de Ciencias de Lima, Union Matematica Argentina and others.

Send subscription order or request for sample copy to

American Mathematical Society, 531 West 116th Street, New York City

TABLE OF CONTENTS

THE CARUS MONOGRAPHS

No. 1. *Calculus of Variations*, by PROFESSOR G. A. BLISS. (First Impression, 1925; Second Impression, 1927; Third Impression, 1935.)

No. 2. *Analytic Functions of a Complex Variable*, by PROFESSOR D. R. CURTISS. (First Impression, 1926; Second Impression, 1930.)

No. 3. *Mathematical Statistics*, by PROFESSOR H. L. RIETZ. (First Impression, 1927; Second Impression, 1929; Third Impression, 1936.)

No. 4. *Projective Geometry*, by PROFESSOR J. W. YOUNG. (First Impression, 1930; Second Impression, 1938.)

No. 5. *History of Mathematics in America before 1900*, by PROFESSORS DAVID EUGENE SMITH AND JEKUTHIEL GINSBURG. (First Impression, 1934.)

No. 6. *Fourier Series and Orthogonal Polynomials*, by PROFESSOR DUNHAM JACKSON. (First Impression, 1941.)

No. 7. *Vectors and Matrices*, by PROFESSOR C. C. MACDUFFEE. (First Impression, 1943.)

Price $1.25 per copy to members of the Mathematical Association, one copy to each member, when ordered directly through the office of the Secretary, MCGRAW HALL, Cornell University, ITHACA, N. Y.

Additional copies for members, and copies for non-members, may be purchased from the Open Court Publishing Co., La Salle, Illinois, at the regular price of $2.00 per copy.

R. R.
8
YALE UNIVERSITY
MAY 13 1944
LIBRARY

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

SOLOMON LEFSCHETZ JOHN VON NEUMANN
FRÉDÉRIC BOHNENBLUST

WITH THE COÖPERATION OF THE
DEPARTMENT OF MATHEMATICS OF PRINCETON UNIVERSITY
AND
THE SCHOOL OF MATHEMATICS OF THE INSTITUTE
FOR ADVANCED STUDY
AND

A. A. ALBERT
EMIL ARTIN
GARRETT BIRKHOFF
M. R. HESTENES[1]
T. H. HILDEBRANDT[2]
NATHAN JACOBSON
N. E. STEENROD
M. H. STONE
ANNA PELL-WHEELER[1]
HASSLER WHITNEY[2]
G. T. WHYBURN[2]
OSCAR ZARISKI

([1] Representing the MATHEMATICAL ASSOCIATION OF AMERICA)
([2] Representing the AMERICAN MATHEMATICAL SOCIETY)

SECOND SERIES, VOL. 45, NO. 2
APRIL, 1944

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE-2, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

COPYRIGHT, 1944, BY PRINCETON UNIVERSITY PRESS

ANNALS OF MATHEMATICS

Subscriptions: The subscription price is \$5.00 a volume (four numbers); the price of a single copy is \$1.50. An arrangement has been made with the MATHEMATICAL ASSOCIATION OF AMERICA whereby an individual member of the ASSOCIATION may subscribe to the ANNALS OF MATHEMATICS at one-half the regular price. Subscription orders, orders for back numbers, notices of change of address, and other business correspondence should be sent to the Princeton University Press, Princeton, N. J.

Editorial Correspondence: Editorial correspondence and manuscripts should be addressed to the ANNALS OF MATHEMATICS, Fine Hall, Princeton, N. J., U. S. A.

Preparation of Manuscripts: Manuscripts submitted to the ANNALS OF MATHEMATICS must be typewritten, double spaced, with generous margins. Figures and diagrams should be drawn in India ink on separate sheets of paper and of twice the size they are to be printed. The lettering is to be omitted from these drawings, but a sketch with complete lettering should be appended.

Footnotes should be numbered consecutively and typed, double spaced, on a sheet by themselves at the end of the manuscript—they will be inserted at the designated places in the text after galley proofs have been corrected. Footnotes should be reduced to a minimum and, whenever possible, replaced by references to a bibliography placed at the end of the paper. Formulae are undesirable in footnotes.

Authors are requested to keep in mind the typographical difficulties involved in setting complicated mathematical formulae. These mechanical difficulties may frequently be eliminated or reduced by the use of alternative notations, *e. g.* "exp" may be used to avoid a complicated exponent, and fractional exponents may be used to replace the troublesome radical sign. The use of unusual symbols is to be avoided. Symbols set one on top of another present a typographical difficulty which usually can be easily avoided by a stepwise arrangement, *e. g.* replace $\sum\limits_{1}^{n}$ by $\sum_1^n$, etc.

Proofs: Authors ordinarily receive only galley proofs. Stylistic changes in proofs must be avoided; authors will be charged for excessive corrections.

Reprints: Authors receive gratis, postage prepaid, one hundred reprints of each article; joint authors each receive fifty reprints.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Made in the United States of America

Copies of the following memoirs can be obtained by addressing *for the United States, Canada, and South America,* The Nordemann Publishing Company, Inc., 215 Fourth Ave., New York, N. Y.

An introduction to the theory of elliptic functions. By GOSTA MITTAG-LEFFLER. Authorized translation by EINAR HILLE. 81 pages. Price 90 cents.

On various conceptions of correlation. By F. M. WEIDA. 37 pages. Price 75 cents.

Contact transformations. By L. P. EISENHART. 40 pages. Price 80 cents.

A survey of the theory of small samples. By P. R. RIDER. 52 pages. Price $1.00.

Tauberian theorems. By NORBERT WIENER. 100 pages. Price $1.00.

Untersuchungen über Gestalt und Lage abgeschlossener Mengen beliebiger Dimension. By PAUL ALEXANDROFF. 86 pages. Price $1.00.

Plateau's problem and Dirichlet's principle. By RICHARD COURANT. 46 pages. Price 90 cents.

Minimal surfaces of higher topological structure. By J. DOUGLAS. 94 pages. Price $1.00.

Sur quelques notions fondamentales dans la théorie générale des operations linéaires. By FRÉDÉRIC RIESZ. 33 pages. Price 70 cents.

La théorie du corps de classes. By C. CHEVALLEY. 25 pages. Price 50 cents.

MATHEMATICAL REVIEWS

offers

prompt review of the mathematical literature of the world
complete subject and author indices in last issue of volume
microfilm or photoprints of most articles reviewed (at cost)

at the very low rate of $13 per year

Publication of this journal is sponsored by the American Mathematical Society, Mathematical Association of America, London Mathematical Society, Academia Nacional de Ciencias de Lima, Union Matematica Argentina and others.

Send subscription order or request for sample copy to

American Mathematical Society, 531 West 116th Street, New York City 27

TABLE OF CONTENTS

THE CARUS MONOGRAPHS

No. 1. *Calculus of Variations*, by PROFESSOR G. A. BLISS. (First Impression, 1925; Second Impression, 1927; Third Impression, 1935.)

No. 2. *Analytic Functions of a Complex Variable*, by PROFESSOR D. R. CURTISS. (First Impression, 1926; Second Impression, 1930.)

No. 3. *Mathematical Statistics*, by PROFESSOR H. L. RIETZ. (First Impression, 1927; Second Impression, 1929; Third Impression, 1936.)

No. 4. *Projective Geometry*, by PROFESSOR J. W. YOUNG. (First Impression, 1930 Second Impression, 1938.)

No. 5. *History of Mathematics in America before 1900*, by PROFESSORS DAVID EUGENE SMITH AND JEKUTHIEL GINSBURG. (First Impression, 1934.)

No. 6. *Fourier Series and Orthogonal Polynomials*, by PROFESSOR DUNHAM JACKSON. (First Impression, 1941.)

No. 7. *Vectors and Matrices*, by PROFESSOR C. C. MACDUFFEE. (First Impression, 1943.)

Price $1.25 per copy to members of the Mathematical Association, one copy to each member, when ordered directly through the office of the Secretary, McGRAW HALL, Cornell University, ITHACA, N. Y.

Additional copies for members, and copies for non-members, may be purchased from the Open Court Publishing Co., La Salle, Illinois, at the regular price of $2.00 per copy.

YALE UNIVERSITY
SEP 19 1944
LIBRARY

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

SOLOMON LEFSCHETZ — JOHN VON NEUMANN

FRÉDÉRIC BOHNENBLUST

WITH THE COÖPERATION OF THE

DEPARTMENT OF MATHEMATICS OF PRINCETON UNIVERSITY

AND

THE SCHOOL OF MATHEMATICS OF THE INSTITUTE
FOR ADVANCED STUDY

AND

A. A. ALBERT
EMIL ARTIN
GARRETT BIRKHOFF
M. R. HESTENES[1]
T. H. HILDEBRANDT[2]
NATHAN JACOBSON
N. E. STEENROD
M. H. STONE
ANNA PELL-WHEELER[1]
HASSLER WHITNEY[2]
G. T. WHYBURN[2]
OSCAR ZARISKI

([1] Representing the MATHEMATICAL ASSOCIATION OF AMERICA)
([2] Representing the AMERICAN MATHEMATICAL SOCIETY)

SECOND SERIES, VOL. 45, NO. 3
JULY, 1944

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE-2, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

COPYRIGHT, 1944, BY PRINCETON UNIVERSITY PRESS

ANNALS OF MATHEMATICS

Subscriptions: The subscription price is \$5.00 a volume (four numbers); the price of a single copy is \$1.50. An arrangement has been made with the MATHEMATICAL ASSOCIATION OF AMERICA whereby an individual member of the ASSOCIATION may subscribe to the ANNALS OF MATHEMATICS at one-half the regular price. Subscription orders, orders for back numbers, notices of change of address, and other business correspondence should be sent to the Princeton University Press, Princeton, N. J.

Editorial Correspondence: Editorial correspondence and manuscripts should be addressed to the ANNALS OF MATHEMATICS, Fine Hall, Princeton, N. J., U. S. A.

Preparation of Manuscripts: Manuscripts submitted to the ANNALS OF MATHEMATICS must be typewritten, double spaced, with generous margins. Figures and diagrams should be drawn in India ink on separate sheets of paper and of twice the size they are to be printed. The lettering is to be omitted from these drawings, but a sketch with complete lettering should be appended.

Footnotes should be numbered consecutively and typed, double spaced, on a sheet by themselves at the end of the manuscript—they will be inserted at the designated places in the text after galley proofs have been corrected. Footnotes should be reduced to a minimum and, whenever possible, replaced by references to a bibliography placed at the end of the paper. Formulae are undesirable in footnotes.

Authors are requested to keep in mind the typographical difficulties involved in setting complicated mathematical formulae. These mechanical difficulties may frequently be eliminated or reduced by the use of alternative notations, *e. g.* "exp" may be used to avoid a complicated exponent, and fractional exponents may be used to replace the troublesome radical sign. The use of unusual symbols is to be avoided. Symbols set one on top of another present a typographical difficulty which usually can be easily avoided by a stepwise arrangement, *e. g.* replace $\sum\limits_{1}^{n}$ by $\sum_1^n$, etc.

Proofs: Authors ordinarily receive only galley proofs. Stylistic changes in proofs must be avoided; authors will be charged for excessive corrections.

Reprints: Authors receive gratis, postage prepaid, one hundred reprints of each article; joint authors each receive fifty reprints.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Made in the United States of America

Copies of the following memoirs can be obtained by addressing *for the United States, Canada, and South America,* The Nordemann Publishing Company, Inc., 215 Fourth Ave., New York, N. Y.

On various conceptions of correlation. By F. M. WEIDA. 37 pages. Price 75 cents.

Contact transformations. By L. P. EISENHART. 40 pages. Price 80 cents.

Tauberian theorems. By NORBERT WIENER. 100 pages. Price $1.00.

Untersuchungen über Gestalt und Lage abgeschlossener Mengen beliebiger Dimension. By PAUL ALEXANDROFF. 86 pages. Price $1.00.

Plateau's problem and Dirichlet's principle. By RICHARD COURANT. 46 pages. Price 90 cents.

Minimal surfaces of higher topological structure. By J. DOUGLAS. 94 pages. Price $1.00.

Sur quelques notions fondamentales dans la théorie générale des operations linéaires. By FRÉDÉRIC RIESZ. 33 pages. Price 70 cents.

La théorie du corps de classes. By C. CHEVALLEY. 25 pages. Price 50 cents.

MATHEMATICAL REVIEWS

offers

prompt review of the mathematical literature of the world

complete subject and author indices in last issue of volume

microfilm or photoprints of most articles reviewed (at cost)

at the very low rate of $13 per year

Publication of this journal is sponsored by the American Mathematical Society, Mathematical Association of America, London Mathematical Society, Edinburgh Mathematical Society, Academia Nacional de Ciencias de Lima, Union Matematica Argentina and others.

Send subscription order or request for sample copy to

American Mathematical Society, 531 West 116th Street, New York City 27

TABLE OF CONTENTS

THE CARUS MONOGRAPHS

No. 1. *Calculus of Variations*, by PROFESSOR G. A. BLISS. (First Impression, 1925; Second Impression, 1927; Third Impression, 1935.)

No. 2. *Analytic Functions of a Complex Variable*, by PROFESSOR D. R. CURTISS. (First Impression, 1926; Second Impression, 1930.)

No. 3. *Mathematical Statistics*, by PROFESSOR H. L. RIETZ. (First Impression, 1927; Second Impression, 1929; Third Impression, 1936.)

No. 4. *Projective Geometry*, by PROFESSOR J. W. YOUNG. (First Impression, 1930 Second Impression, 1938.)

No. 5. *History of Mathematics in America before 1900*, by PROFESSORS DAVID EUGENE SMITH AND JEKUTHIEL GINSBURG. (First Impression, 1934.)

No. 6. *Fourier Series and Orthogonal Polynomials*, by PROFESSOR DUNHAM JACKSON. (First Impression, 1941.)

No. 7. *Vectors and Matrices*, by PROFESSOR C. C. MACDUFFEE. (First Impression, 1943.)

Price $1.25 per copy to members of the Mathematical Association, one copy to each member, when ordered directly through the office of the Secretary, MCGRAW HALL, Cornell University, ITHACA, N. Y.

Additional copies for members, and copies for non-members, may be purchased from the Open Court Publishing Co., La Salle, Illinois, at the regular price of $2.00 per copy.

YALE UNIVERSITY
NOV 17 1944
LIBRARY

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

SOLOMON LEFSCHETZ JOHN VON NEUMANN

FRÉDÉRIC BOHNENBLUST

WITH THE COÖPERATION OF THE

DEPARTMENT OF MATHEMATICS OF PRINCETON UNIVERSITY

AND

THE SCHOOL OF MATHEMATICS OF THE INSTITUTE FOR ADVANCED STUDY

AND

A. A. ALBERT
EMIL ARTIN
GARRETT BIRKHOFF
M. R. HESTENES[1]
T. H. HILDEBRANDT[2]
NATHAN JACOBSON
N. E. STEENROD
M. H. STONE
ANNA PELL-WHEELER[1]
HASSLER WHITNEY[2]
G. T. WHYBURN[2]
OSCAR ZARISKI

([1] Representing the MATHEMATICAL ASSOCIATION OF AMERICA)
([2] Representing the AMERICAN MATHEMATICAL SOCIETY)

SECOND SERIES, VOL. 45, NO. 4
OCTOBER, 1944

PUBLISHED QUARTERLY
AT
MOUNT ROYAL AND GUILFORD AVENUES
BALTIMORE-2, MD.
BY THE
PRINCETON UNIVERSITY PRESS
PRINCETON, N. J.

COPYRIGHT, 1944, BY PRINCETON UNIVERSITY PRESS

ANNALS OF MATHEMATICS

Subscriptions: The subscription price is \$5.00 a volume (four numbers); the price of a single copy is \$1.50. An arrangement has been made with the MATHEMATICAL ASSOCIATION OF AMERICA whereby an individual member of the ASSOCIATION may subscribe to the ANNALS OF MATHEMATICS at one-half the regular price. Subscription orders, orders for back numbers, notices of change of address, and other business correspondence should be sent to the Princeton University Press, Princeton, N. J.

Editorial Correspondence: Editorial correspondence and manuscripts should be addressed to the ANNALS OF MATHEMATICS, Fine Hall, Princeton, N. J., U. S. A.

Preparation of Manuscripts: Manuscripts submitted to the ANNALS OF MATHEMATICS must be typewritten, double spaced, with generous margins. Figures and diagrams should be drawn in India ink on separate sheets of paper and of twice the size they are to be printed. The lettering is to be omitted from these drawings, but a sketch with complete lettering should be appended.

Footnotes should be numbered consecutively and typed, double spaced, on a sheet by themselves at the end of the manuscript—they will be inserted at the designated places in the text after galley proofs have been corrected. Footnotes should be reduced to a minimum and, whenever possible, replaced by references to a bibliography placed at the end of the paper. Formulae are undesirable in footnotes.

Authors are requested to keep in mind the typographical difficulties involved in setting complicated mathematical formulae. These mechanical difficulties may frequently be eliminated or reduced by the use of alternative notations, *e. g.* "exp" may be used to avoid a complicated exponent, and fractional exponents may be used to replace the troublesome radical sign. The use of unusual symbols is to be avoided. Symbols set one on top of another present a typographical difficulty which usually can be easily avoided by a stepwise arrangement, *e. g.* replace $\sum\limits_{1}^{n}$ by $\sum_1^n$, etc.

Proofs: Authors ordinarily receive only galley proofs. Stylistic changes in proofs must be avoided; authors will be charged for excessive corrections.

Reprints: Authors receive gratis, postage prepaid, one hundred reprints of each article; joint authors each receive fifty reprints.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879.

Made in the United States of America

Copies of the following memoirs can be obtained by addressing

for the United States, Canada, and South America, The Nordemann Publishing Company, Inc., 215 Fourth Ave., New York, N. Y.

On various conceptions of correlation. By F. M. WEIDA. 37 pages. Price 75 cents.

Contact transformations. By L. P. EISENHART. 40 pages. Price 80 cents.

Tauberian theorems. By NORBERT WIENER. 100 pages. Price $1.00.

Untersuchungen über Gestalt und Lage abgeschlossener Mengen beliebiger Dimension. By PAUL ALEXANDROFF. 86 pages. Price $1.00.

Plateau's problem and Dirichlet's principle. By RICHARD COURANT. 46 pages. Price 90 cents.

Minimal surfaces of higher topological structure. By J. DOUGLAS. 94 pages. Price $1.00.

Sur quelques notions fondamentales dans la théorie générale des operations linéaires. By FRÉDÉRIC RIESZ. 33 pages. Price 70 cents.

La théorie du corps de classes. By C. CHEVALLEY. 25 pages. Price 50 cents.

MATHEMATICAL REVIEWS

offers

prompt review of the mathematical literature of the world

complete subject and author indices in last issue of volume

microfilm or photoprints of most articles reviewed (at cost)

at the very low rate of $13 per year

Publication of this journal is sponsored by the American Mathematical Society, Mathematical Association of America, London Mathematical Society, Edinburgh Mathematical Society, Academia Nacional de Ciencias de Lima, Union Matematica Argentina and others.

Send subscription order or request for sample copy to

American Mathematical Society, 531 West 116th Street, New York City 27

TABLE OF CONTENTS

Outline of the History of Mathematics

by RAYMOND CLARE ARCHIBALD

Fifth edition, June 1941, ii, 76 pages

THIS thoroughly revised and considerably enlarged edition is published by the Mathematical Association to meet a constant demand for up-to-date information on the History of Mathematics, not available in any other single English work. The syllabus, twenty pages of references to other material and sources, and an entirely new Index of Names, provide an excellent basis for a teacher to conduct a semester or year course in this field, or for a student wishing greatly to extend his knowledge. Earlier editions of the Outline have been reviewed very favorably throughout the world. Our expectation is that this fifth edition published within ten years, and more fundamentally revised and extended than any other edition, will continue to meet a need in this country and elsewhere.

Price 75 cents a copy, postpaid, remittance with order

No discount in price to anyone

MATHEMATICAL ASSOCIATION OF AMERICA
McGRAW HALL, CORNELL UNIVERSITY
ITHACA, NEW YORK

3